

Wenn aus Daten Journalismus wird: eine Inhaltsanalyse der für die Data Journalism Awards 2013 bis 2015 nominierten Projekte

Loosen, Wiebke; Reimer, Julius; Silva-Schmidt, Fenja de

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Loosen, W., Reimer, J., & Silva-Schmidt, F. d. (2016). *Wenn aus Daten Journalismus wird: eine Inhaltsanalyse der für die Data Journalism Awards 2013 bis 2015 nominierten Projekte*. (Arbeitspapiere des Hans-Bredow-Instituts, 39). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI). <https://doi.org/10.21241/ssoar.71789>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>



HANS-BREDOW-INSTITUT
für Medienforschung an der Universität Hamburg

Wiebke Loosen / Julius Reimer / Fenja de Silva-Schmidt

Wenn aus Daten Journalismus wird

Eine Inhaltsanalyse der für die Data Journalism Awards
2013 bis 2015 nominierten Projekte

Oktober 2016

Arbeitspapiere des Hans-Bredow-Instituts Nr. 39

Loosen, Wiebke; Reimer, Julius; de Silva-Schmidt, Fenja (2016): Wenn aus Daten Journalismus wird. Eine Inhaltsanalyse der für die Data Journalism Awards 2013 bis 2015 nominierten Projekte. Hamburg: Verlag Hans-Bredow-Institut, Oktober 2016. DOI: <https://doi.org/10.21241/ssoar.71789>.

ISSN 1435-9413

ISBN 978-3-87296-137-2

Die Hefte der Schriftenreihe „Arbeitspapiere des Hans-Bredow-Instituts“ finden sich zum Download auf der Website des Instituts unter der Adresse <http://hbi.to/957>.

Ein Ausdruck des Heftes ist gegen eine Schutzgebühr von 20,00 EUR direkt beim Verlag erhältlich.

Hans-Bredow-Institut für Medienforschung an der Universität Hamburg

Medienvermittelte öffentliche Kommunikation – das ist das Forschungsgebiet des Hans-Bredow-Instituts für Medienforschung. Die Medien prägen heute mehr denn je den Alltag, die Politik, die Wirtschaft und die Kultur. Kernziel der Forschung des Instituts ist es, diese Einflüsse der Medien zu verstehen, Entwicklungen und Risiken abzuschätzen und Handlungsoptionen zu entwickeln. Das Institut wahrt dabei seine unabhängige Position. Die Wissenschaftler sehen ihre Verantwortung darin, sich neuen Fragestellungen nach wissenschaftlichen Kriterien zu nähern, ihre Grundannahmen und Methoden transparent zu machen. Die Erforschung der aktuellen Medienentwicklung erfolgt interdisziplinär. Die für das Institut grundlegenden kommunikations- und rechtswissenschaftlichen Perspektiven werden mit weiteren Ansätzen, etwa aus Ökonomie und Pädagogik, kombiniert. Eine derartige Medienforschung setzt Kooperation voraus. Mit Partnern in vielen Ländern werden international vergleichende Fragestellungen bearbeitet. Das Institut pflegt vielfältige Kontakte und sucht den steten Austausch mit den verschiedenen Akteuren. Dazu gehört der Transfer von Forschungsergebnissen in die Praxis, Politik und Öffentlichkeit hinein. Mehr unter www.hans-bredow-institut.de.

Die Autoren

Julius Reimer, M.A., ist Junior Researcher, PD Dr. Wiebke Loosen ist Senior Researcher am Hans-Bredow-Institut für Medienforschung in Hamburg, Fenja de Silva-Schmidt arbeitet im Institut für Journalistik und Kommunikationswissenschaft der Universität Hamburg.

Projekt „Wenn aus Daten Journalismus wird“

Ob „Swiss Leaks“ oder „Panama Papers“ – „Datenjournalismus“ ist derzeit hoch aktuell. Wie sich diese neue Form von Journalismus entwickelt, wird seit 2013 anhand journalistischer Texte untersucht, die für den Data Journalism Award nominiert wurden. Dieser Preis wird jedes Jahr vom Global Editors Network für die besten datengetriebenen Geschichten vergeben. Analysiert wird, welche Medien die Stücke produzieren (Zeitungen, TV-Sender usw.), welche Themen sie behandeln (Politik, Wirtschaft, Sport usw.), welche Arten von Daten sie dafür nutzen (Geodaten, Finanzdaten, Umfrageergebnisse usw.), aus welchen Quellen diese stammen (Regierungsorganisationen, NGOs, Unternehmen usw.), mit welchen Mitteln sie visuell dargestellt werden (Diagramme, Karten, Tabellen usw.) und welche Möglichkeiten sie den Nutzerinnen und Nutzern bieten, die Daten selbst interaktiv zu erkunden (in Karten hineinzoomen, Daten nach dem eigenen Wohnort filtern usw.). Durch den Vergleich über die einzelnen Jahre werden Trends in der Entwicklung des noch jungen Berichterstattungsstils ebenso deutlich wie bisher ungenutzte Potenziale. Weitere Informationen unter <http://hbi.to/6548>.

Hans-Bredow-Institut für Medienforschung an der Universität Hamburg

Verlag

Rothenbaumchaussee 36

20148 Hamburg

Tel.: (+49 40) 450 217-12

Fax: (+49 40) 450 217-77

E-Mail: info@hans-bredow-institut.de

Inhalt

Zusammenfassung	5
1. Einleitung und Forschungsstand.....	7
2. Methode.....	11
2.1 Auswahl des Analysematerials	11
2.2 Codebuch.....	12
3. Ergebnisse	15
3.1 Medientypen und Anzahl der beteiligten Akteure.....	15
3.2 In den datenjournalistischen Projekten behandelte Themen	18
3.3 Erzählerische Mittel und Elemente von Kritik.....	20
3.4 Datentypen und -verarbeitung.....	21
3.5 Verwendete Formen der Visualisierung.....	27
3.6 Interaktive Funktionen	30
4. Zusammenfassung und Fazit.....	32
Literatur	37

Zusammenfassung

Die Entstehung des Datenjournalismus kann als Antwort des Journalismus auf die „Datafizierung“ der Gesellschaft verstanden werden: In der Auseinandersetzung mit dem Phänomen „Big Data“ entwickelt der Journalismus neue Wege, Geschichten (in Daten) zu identifizieren und (mit Daten) zu erzählen. Von Journalismusforschern wie -praktikern wird das dabei entstehende Berichterstattungsmuster häufig als *die* Zukunft des Journalismus angesehen, vor allem aber als genuin für den Onlinejournalismus. Aus Sicht der Forschung ist jedoch noch nicht einmal der Status Quo des ohnehin sehr dynamischen Datenjournalismus geklärt: Noch wissen wir wenig insbesondere über das, was den Datenjournalismus als Berichterstattungsmuster ausmacht, nämlich die datenjournalistischen Beiträge sowie die spezifischen inhaltlichen und darstellerischen Elemente, aus denen sie bestehen und die sie von anderen Formen der Berichterstattung abgrenzen. Der Schwerpunkt dieser Studie liegt daher auf den Produkten des Datenjournalismus: Mithilfe einer standardisierten Inhaltsanalyse untersuchen wir datenjournalistische Projekte, die man als „Goldstandard“ datengetriebener Berichterstattung betrachten kann – Projekte, die von 2013 bis 2015 (N = 179) für einen der *Data Journalism Awards* (DJA) nominiert wurden, einen jährlich vom Global Editors Network in unterschiedlichen Kategorien vergebenen Preis. Untersucht wurden unter anderem die genutzten Datenarten und -quellen, Visualisierungsstrategien, Interaktionsmöglichkeiten, die behandelten Themen sowie die Medienangebote, von denen die Beiträge stammen. Die Ergebnisse zeigen unter anderem, dass über 40 Prozent der datengetriebenen Arbeiten auf Webseiten von Tages- oder Wochenzeitungen veröffentlicht wurden und fast 20 Prozent von Non-Profit Organisationen für investigativen Journalismus wie *Pro Publica* stammen. Fast die Hälfte der Fälle behandelt ein politisches Thema. Auch Gesellschafts- und Wirtschafts- sowie Gesundheits- und Wissenschaftsthemen kommen vermehrt vor. Finanz- und Geodaten sind die meistgenutzten Datenarten, wobei sich der Großteil der Datensätze auf eine nationale Ebene bezieht. In mehr als zwei Drittel der Projekte werden die Daten einer offiziellen Quelle wie zum Beispiel *Eurostat* verwendet. In weiteren Analysen werden die Unterschiede zwischen den Jahren 2013, 2014 und 2015 herausgearbeitet und Visualisierungsstrategien sowie Interaktionsmöglichkeiten untersucht.

1. Einleitung und Forschungsstand¹

Wie „Big Data“ und die „Macht der Algorithmen“ (Beer 2009: 985ff.) die Gesellschaft verändern, gehört sicher zu den spannendsten und wichtigsten Fragen unserer Zeit. Kein gesellschaftlicher Bereich scheint davon unberührt zu bleiben: „Big Data“ sind Thema unter anderem in der Politik, Wirtschaft, Wissenschaft und Medizin (vgl. Mayer-Schönberger/Cukier 2013).

Für den Journalismus ist eine „datengetriebene Gesellschaft“ in doppelter Hinsicht relevant: Zum einen sind damit Entwicklungen angesprochen, die Gegenstand (kritischer) Berichterstattung werden (müssen), um einen öffentlichen Diskurs zu ermöglichen über Datensicherheit, Datenschutz und andere Folgen und Rahmenbedingungen von „Datafizierung“ (vgl. Borges-Rey 2016). Zum anderen beobachten wir im Journalismus selbst einen „quantitative turn“ (Coddington 2015), in dessen Zuge Daten mehr und mehr zum „Rohstoff“ für ein Berichterstattungsmuster (vgl. Schmidt/Weischenberg 1994) „Datenjournalismus“ werden. Das Ergebnis ist die Entstehung eines neuen journalistischen Genres, das mit Bezeichnungen wie „computational or algorithmic journalism“ (Anderson 2013), „data journalism“ (Appelgren/Nygren 2014) oder „data-driven journalism (Parasie/Dagiral 2013) beschrieben wird.

Die Entstehung des Datenjournalismus kann als Antwort des Journalismus auf die „Datafizierung“ (Filipović 2015) der Gesellschaft und den dadurch erzeugten „data deluge“ (Lewis 2015: 322) verstanden werden. Tatsächlich bringt die Auseinandersetzung des Journalismus mit dem Phänomen „Big Data“ neuartige Möglichkeiten hervor, journalistische Geschichten (in Daten) zu entdecken und (mit Daten) zu erzählen (vgl. Anderson 2013; Coddington 2015; Lewis/Usher 2014). Die intensive Auseinandersetzung mit dem Phänomen „Datenjournalismus“ im Journalismus selbst (vgl. z. B. Fink/Anderson 2015; Weinacht/Spiller 2014) verdeutlicht die hohe praktische Relevanz des Themas und seine Fähigkeit, journalistische Selbstreflexion anzuregen. Gleichzeitig beeinflusst der Trend bereits die journalistische Ausbildung: Diverse Universitäten bieten inzwischen Datenjournalismus-Seminare oder gar -Studiengänge an (vgl. Anderson 2013; Splendore et al. 2016; Weinacht/Spiller 2014).

Die weitreichende Aufmerksamkeit, die dem Datenjournalismus von Seiten der Praktiker zuteilwird, zieht eine „explosion in data journalism-oriented scholarship“ (Fink/Anderson 2015: 467) nach sich. Dieser „rapidly growing body“ (Lewis, 2015: 322) an wissenschaftlichen Studien fokussiert dabei bisher im Wesentlichen auf drei Aspekte: Erstens wurden die Akteure im Datenjournalismus untersucht, zweitens sich häufig widersprechende Definitionen des Datenjournalismus aufgestellt und drittens datenjournalistische Beiträge analysiert.

Für den ersten Forschungsstrang stehen die Befragungen von Datenjournalisten in Belgien (De Maeyer et al. 2015), Deutschland (Weinacht/Spiller 2014), Kanada (Tabary et al. 2016),

¹ Für ihre Unterstützung bei der Erstellung dieses Arbeitspapiers bedanken wir uns herzlich bei Hannah Immler.

Norwegen (Karlsen/Stavelin 2014), Schweden (Appelgren/Nygren 2014), dem Vereinigten Königreich (Borges-Rey 2016) und den Vereinigten Staaten (Boyles/Meyer 2016; Fink/Anderson 2015; Parasie 2015; Parasie/Dagiral 2013), in denen es u. a. um das Rollenselbstverständnis der Befragten sowie ihre organisationale Einbettung in Redaktionen geht. Diese Studien weisen darauf hin, dass datengetriebene Projekte nur von einer „handful of consensual actors“ (De Maeyer et al. 2015: 438), d. h. von einigen wenigen Experten umgesetzt werden (vgl. auch Appelgren/Nygren 2014; Karlsen/Stavelin 2014; Weinacht/Spiller 2014). So verfügen oft nur große, überregional agierende Medienorganisationen wie etwa *The Guardian*, *The New York Times* oder *Pro Publica* über genügend Ressourcen, um datenjournalistische „Leuchtturmprojekte“ (Weinacht/Spiller 2014: 428) zu realisieren (vgl. De Maeyer et al. 2015; Karlsen/Stavelin 2014; Parasie/Dagiral 2013). Dementsprechend werden Ressourcenknappheit, Zeit- und Personalmangel als vornehmliche Hindernisse für eine Ausweitung des Datenjournalismus insbesondere in kleinen Redaktionen wahrgenommen (vgl. Appelgren/Nygren 2014; Boyles/Meyer 2016; De Maeyer et al. 2015; Fink/Anderson 2015; Parasie 2015; Weinacht/Spiller 2014). Trotzdem messen die Befragten in mehreren Studien in den untersuchten Ländern datengetriebenen Projekten eine wachsende Bedeutung bei und prognostizieren einen Aufwärtstrend für die Zukunft (vgl. Appelgren/Nygren 2014; De Maeyer et al. 2015; Parasie/Dagiral 2013; Weinacht/Spiller 2014).

Auch Gemeinsamkeiten hinsichtlich der Arbeitsweise in datengetriebenen Projekten konnten bereits ausgemacht werden: Oft werden sie in Teams realisiert, in denen Spezialisten aus den Bereichen Recherche und journalistische Darstellung, Grafik und Visualisierung sowie Programmierung und Datenanalyse zusammenarbeiten (vgl. Boyles/Meyer 2016; Karlsen/Stavelin 2014; Parasie 2015; Parasie/Dagiral 2013; Weinacht/Spiller 2014; Tabary et al. 2016). De Maeyer et al. (2015: 440f.) differenzieren in diesem Zusammenhang zwischen „ordinary data journalism“ – kleineren Geschichten, die in Einzelarbeit produziert werden – und „thorough data journalism“ – komplexeren Projekten, bei denen ein Team mit vielfältigen Kompetenzen gefordert ist. Die befragten Journalisten haben daher häufig eine Vorbildung aus dem Technik- oder Grafikbereich oder Weiterbildungen in diesen Richtungen absolviert (vgl. Karlsen/Stavelin 2014; Fink/Andersen 2015; Parasie/Dagiral 2013). Ihre Aufgaben in der Redaktion sind zudem nicht immer eindeutig als „Datenjournalismus“ bezeichnet und nicht trennscharf von anderen Arbeitsbereichen abgegrenzt (vgl. Fink/Andersen 2015; Parasie/Dagiral 2013; Weinacht/Spiller 2014; Tabary et al. 2016).

Gefragt nach ihrem Rollenselbstverständnis als Journalisten heben die Befragten in den jeweiligen Ländern vor allem die Kontroll- und Kritikfunktion hervor (vgl. Weinacht/Spiller 2014) samt der Aufgabe, Missstände aufzuzeigen (vgl. Parasie/Dagiral 2013) etwa mittels investigativer Recherche und schwer zugänglichen oder gar „geleakten“ Datensätzen (vgl. De Maeyer et al. 2015; Karlsen/Stavelin 2014). Einige Befragte stellen Crowdsourcing und die Interaktion mit Nutzern zum Zwecke der Datenbeschaffung und -verifikation als wichtige Bestandteile datenjournalistischer Arbeit heraus:

„The potential of using human judgment on the Web to gather and verify information is substantial, and it is possible to see how that can transform investigative journalism in some cases“. (Karlsen/Stavelin 2014: 44; vgl. auch Appelgren/Nygren 2014; Borges-Rey 2016; Boyles/Meyer 2016)

Den zweiten Strang wissenschaftlicher Bemühungen stellen Versuche dar, herauszuarbeiten, was Datenjournalismus im Kern ausmacht und wo Gemeinsamkeiten und Unterschiede im Vergleich zu verwandten Ansätzen wie dem Computer-assisted Reporting (CAR), computational journalism oder „herkömmlicher“ investigativer Recherche liegen. Diese Klassifizierungsversuche z. B. von Coddington (2015), Fink/Anderson (2015) oder auch Royal/Blasingame (2015) widersprechen sich jedoch nicht selten. Während zum Beispiel Anderson (2013: 1005) datengetriebenen Journalismus als Teil des weiteren Felds des „computational journalism“ versteht, unterscheidet Coddington (2015) explizit zwischen diesen beiden Konzepten. Dementsprechend beklagen Fink/Anderson (2015: 478) auch das Fehlen einer „shared definition of data journalism“. Basierend auf kursorischen Beobachtungen des Feldes und Fallanalysen von Beispielpunkten werden dennoch immer wieder folgende Schlüsselemente des Datenjournalismus identifiziert: Datenjournalismus

- basiert auf in der Regel großen Mengen quantitativer Daten als „Rohmaterial“, das (statistisch) ausgewertet wird, um Geschichten von journalistischer Relevanz zu entdecken und zu erzählen (vgl. Coddington 2015; Royal/Blasingame 2015);
- ist charakterisiert durch „participatory openness“ (Coddington 2015: 337) und „crowdsourcing“ (Appelgren/Nygren 2014: 394), insofern als Nutzer mithelfen, Daten zu sammeln, zu analysieren und/oder zu interpretieren;
- wird regelmäßig in Verbindung gebracht mit einem Open data- und Open source-Ansatz, nach dem es etwa als Qualitätskriterium gilt, dass auch die Rohdaten veröffentlicht werden, auf denen eine Geschichte basiert (vgl. Gray et al. 2012);
- seine Ergebnisse bedürfen oft der Visualisierung (vgl. Gray et al. 2012: o. S.), damit sie allgemeinverständlich werden, d. h. sie werden in Form von Karten, Balkendiagrammen und anderen Grafiken aufbereitet und präsentiert (vgl. Royal/Blasingame 2015).

Der dritte Forschungsstrang schließlich beschäftigt sich mit Inhalten, die als datenjournalistisch identifiziert werden, setzt also auf der Ebene der Produkte an. In diesem Bereich ist bisher nur eine Handvoll von Arbeiten erschienen, die ebenfalls auf die oben genannten Charakteristika fokussieren und damit ihren Status als „Schlüsselemente“ eines datenjournalistischen Berichterstattungsmusters bestätigen. Allerdings sind die analysierten Stichproben mehr oder weniger stark raum-zeitlich eingeschränkt: Parasie/Dagiral (2013) untersuchen Stücke lediglich einer Chicagoer Zeitung, die vor März 2011 veröffentlicht wurden; Knight (2015) analysiert Artikel aus 15 britischen Zeitungen, die allerdings in einem Zeitraum von nur zwei Wochen erschienen sind; Tandoc/Oh (2015) wenden sich 260 Beiträgen aus den Jahren 2009 bis 2015 zu, die jedoch sämtlich aus dem *Datablog* des *Guardian* stammen; und Tabary et al. (2016) befassen sich mit Projekten aus den Jahren 2011 bis 2013, die aber alle von nur sechs Medien aus Quebec verantwortet werden.

Die bisher vorliegende Forschung zum Datenjournalismus liefert somit:

- Erkenntnisse zu den Akteuren im Datenjournalismus, ihrem journalistischen Selbstverständnis sowie ihrer Einbettung in und Anbindung an Redaktionen,
- mehrere Versuche, Datenjournalismus als abgrenzbares Berichterstattungsmuster zu definieren, das sich um eine Reihe von Kernelementen dreht,
- erste Fallstudien, die empirisch datenjournalistische Projekte untersuchen, jedoch in Umfang und Aussagekraft begrenzt sind.

Was fehlt ist eine Längsschnittstudie, die uns hilft, zu verstehen, wie sich Datenjournalismus als „*emerging form of storytelling*“ (Appelgren/Nygren 2014: 394; Hervorh. d. Verf.) über die Zeit (weiter-)entwickelt, das heißt, ob die Kerncharakteristika des Datenjournalismus in neuen Formen auftreten oder sich die Liste dieser Schlüsselemente selbst ändert. Dieses Defizit führt zu der paradoxen Situation, dass Datenjournalismus von Journalisten wie Forschern regelmäßig als zentral für die *Zukunft* des Journalismus beschrieben wird (vgl. Gray et al. 2012; Lewis/Usher 2014), obwohl nicht klar ist, wie er sich *gegenwärtig* entwickelt.

Die vorliegende Studie hat daher zum Ziel, das Verständnis der fortlaufenden (R)Evolution der datengetriebenen Berichterstattung zu fördern, indem eine Längsschnittperspektive eingenommen wird. Dies geschieht mit einem vornehmlich deskriptiven Ansatz, der für die Erforschung eines vergleichsweise neuen und sich beständig entwickelnden Phänomens wie diesem angemessen erscheint, da er nicht durch eine spezifische theoretische Brille eingeschränkt ist und so das empirische Fundament für weitere Analysen und ein genaueres Verständnis des Phänomens „Datenjournalismus“ legen kann.

Das Vorgehen folgt folgenden Forschungsfragen:

FF 1: Aus welchen strukturellen und darstellerischen Elementen bestehen datengetriebene journalistische Stücke und wie entwickeln sie sich über die Zeit?

FF 2: Welche Themen werden datenjournalistisch behandelt und wie verändert sich die thematische Ausrichtung des Berichterstattungsmusters über die Zeit?

FF 3: Wie entwickelt sich das Feld der Akteure, also der Medienorganisationen, redaktionellen Teams und externer Partner, die Datenjournalismus produzieren?

2. Methode

Zur Beantwortung der genannten Forschungsfragen führen wir eine standardisierte Inhaltsanalyse (vgl. Krippendorff 2013) datenjournalistischer Projekte durch, wobei die Stichprobe räumlich und zeitlich breiter angelegt ist als bei den oben genannten bereits vorliegenden Studien, dafür aber andere Restriktionen in Kauf nehmen muss.

2.1 *Auswahl des Analysematerials*

Der Forschungsstand zeigt, dass Datenjournalismus weiterhin ein „diffuse term“ (Fink/Anderson 2015: 470) bleibt. Dies macht die Auswahl relevanter datenjournalistischer Produkte für die Inhaltsanalyse schwierig, zumindest aber voraussetzungsreich. Daher wählen wir ein pragmatisches und induktives Verfahren, durch das gewährleistet werden soll, dass der Stichprobe weder eine zu eingeschränkte noch eine zu weite Definition von Datenjournalismus zugrunde liegt. Die Auswahl stützt sich zu diesem Zweck auf eine Definition aus dem Feld selbst. Datengrundlage sind Projekte, die in den Jahren 2013, 2014 und 2015 für einen der Data Journalism Awards (DJA) nominiert wurden. Diese Auszeichnung wird seit 2012 jährlich vom Global Editors Network (GEN) in mehreren Kategorien verliehen, unterstützt von Google und der Knight Foundation.² Damit unterliegt das Sample einem doppelten Selektionsbias: Zunächst ist es selbstselektiv, da Journalisten ihre Beiträge selbst einreichen müssen, um überhaupt nominiert werden zu können. Im zweiten Schritt entscheidet dann eine in der Besetzung jährlich mehr oder weniger wechselnde Fachjury, welche Einreichungen tatsächlich nominiert werden. Die konstant hohen Einreichungszahlen (s. Tabelle 1) legen jedoch nahe, dass trotz Selbstselektion ein großer Teil der (besonders elaborierten) datenjournalistischen Veröffentlichungen der drei Jahre vertreten ist. Der zweite Selektionsschritt wiederum ist erwünscht, da durch ihn garantiert wird, dass wir Projekte analysieren, die innerhalb des Feldes selbst als Datenjournalismus angesehen werden und als bedeutsame oder innovative Beispiele ausgewählt wurden. Auch Weinacht/Spiller (2014: 412) gehen davon aus, dass die Vorstellung von diesem noch neuen Phänomen geschärft werden kann, „wenn herausragende Beispiele betrachtet werden“. Zudem ist davon auszugehen, dass diese besonderen Projekte eine gewisse Signalwirkung für das Feld insgesamt haben und unter Umständen als Vorbild für folgende datenjournalistische Projekte dienen. Vergleichbare Auswahlverfahren haben sich bereits bei der Untersuchung anderer Formen journalistischer Berichterstattung bewährt; Wahl-Jorgensen (2013a, 2013b) und Lanosga (2014) haben z. B. anhand der Nominierten und Gewinner des Pulitzer-Preises, Emotionalität im Journalismus beziehungsweise investigative Berichterstattung untersucht.

Tabelle 1 gibt einen Überblick über die Stichprobe. Nominierungen, die sich auf eine Redaktion oder ein Medienunternehmen als Ganzes beziehen und nicht auf ein einzelnes Pro-

² Vgl. <http://www.globaleditorsnetwork.org/programmes/data-journalism-awards/> [24.10.2016].

jekt, wurden von der Analyse ausgeschlossen, da die Analyseeinheit das einzelne datenjournalistische Projekt/Produkt ist. In mehreren DJA-Kategorien nominierte Beiträge wurden jeweils nur einmal gezählt.

Tabelle 1: Zusammensetzung der Stichprobe

		2013	2014	2015	Gesamt
Einreichungen	<i>Häufigkeit</i>	>300 ¹	520	482	>1.302
Nominierte Projekte	<i>Häufigkeit</i>	72	75	78	225
Für die Analyse geeignete Projekte	<i>Häufigkeit</i>	56	64	59	179
	<i>% der nominierten Projekte des jeweiligen Jahres</i>	77,8	85,3	75,6	79,6
	<i>% der analysierten Projekte</i>	31,3	35,8	33,0	100,0
Ausgezeichnete Projekte	<i>Häufigkeit</i>	6	9	13	28
	<i>% der analysierten Projekte des jeweiligen Jahres</i>	10,7	14,1	22,0	15,6

¹ Das GEN macht keine präzise Angabe über die Anzahl der Einreichungen für das Jahr 2013, erklärt jedoch, dass „more than 300 entries“ gemacht wurden (<http://www.globaleditorsnetwork.org/programmes/dja> [24.02.2014]).

Im Hinblick auf die oben formulierten Forschungsfragen ermöglicht diese Stichprobe, Unterschiede zwischen den Jahren 2013, 2014 und 2015 zu identifizieren und somit die Entwicklung des datenjournalistischen Berichterstattungsmusters (unter den gegebenen Einschränkungen des Samples) nachzuzeichnen. Darüber hinaus erlaubt das Material, Differenzen zwischen den Gewinnern eines DJA und jenen datenjournalistischen Arbeiten, die lediglich nominiert waren, auszumachen.

2.2 Codebuch

Das Codebuch für die Inhaltsanalyse umfasst die oben diskutierten (mutmaßlichen) „Schlüsselemente“ des Datenjournalismus sowie weitere (Kontext-)Variablen. Die meisten Variablen und Kategorien wurden induktiv auf Basis einer explorativen Analyse von 56 DJA-nominierten Beiträgen aus dem Jahr 2013 entwickelt. Einige Variablen wurden von der bereits genannten Studie von Parasie/Dagiral (2013) inspiriert, andere gehen auf Anregungen des (Daten-) Journalismusforschers Julian Ausserhofer und des Datenjournalisten Lorenz Matzat zurück. Bei einem Pretest mit zwei Codierern und einem Subsample von zehn Prozent der Fälle aus den Jahren 2013 und 2014 ($n = 12$) erreichten alle Variablen eine InterCodererreliabilität (Holsti bzw. – wo möglich – Krippendorffs Alpha) von 0,7 oder höher, was für eine explorative Studie als ausreichend erachtet wird (vgl. Lombard et al. 2002). Das endgültige Codebuch enthält 29 Variablen, die sich grob fünf Oberkategorien zuordnen lassen (s. Tabelle 2). Im vorliegenden Arbeitspapier stellen wir eine Auswahl der analysierten Variablen vor.

Tabelle 2: Dimensionen und Variablen des Codebuchs³

Dimension	Variablen
Autoren	<ul style="list-style-type: none"> - Medium - Art des Mediums - Herkunftsland des Mediums - externe Partner - Anzahl der namentlich genannten Beteiligten
Projekteigenschaften	<ul style="list-style-type: none"> - Titel des nominierten Projekts - Thema des nominierten Projekts - Sprache - DJA-Gewinner
Datensatz	<ul style="list-style-type: none"> - Datenquelle(n) - Art der Datenquelle(n) - Zugang zu den Daten - Art der Daten - Zeitlicher Bezug der Daten^I - Geografischer Bezug der Daten - Veränderbarkeit der Daten^{II} - Abgedeckte Zeitspanne - Analyseeinheit - Zusätzliche Informationen zu den Daten^I
Analyse und erzählerisch-darstellerische Mittel	<ul style="list-style-type: none"> - An die Daten gestellte Frage - Fokus der Analyse (Darstellung von Häufigkeiten/Zeitverläufen/Hierarchien)^{IV} - Zugehörige Textbeiträge^{III} - Länge der zugehörigen Beiträge - Personalisiertes Fallbeispiel^V - Kritik an wahrgenommenen Missständen und/oder Aufruf an Nutzer, Veränderungen einzufordern, bzw. direkte Forderung nach einem Eingreifen der Politik^{II} - Visualisierungen
Nutzungskontext	<ul style="list-style-type: none"> - Interaktive Funktionen - Online-Zugang zur Datenbank^{III} - Datenbezogene Anschlusskommunikation/-handlung

^I Vorgeschlagen von Datenjournalist Lorenz Matzat^{II} Vorgeschlagen von (Daten-)Journalismusforscher Julian Ausserhofer^{III} Übernommen von Parasie und Dagiral (2013: 5–14)^{IV} In Anlehnung an Gray et al. (2012: o.S.)^V In Anlehnung an Holtermann (2011)

Über die drei Jahre hinweg mussten lediglich geringfügige Änderungen am Codebuch vorgenommen werden: So wurden etwa bei der Variable „Herkunftsland des Mediums“ weitere Länder als Ausprägungen hinzugefügt; im Jahr 2015 wurden zum ersten Mal Audio-Dateien als Darstellungsmittel in einem Projekt genutzt, sodass eine entsprechende Ausprägung ergänzt werden musste. Ansonsten ließen sich alle über die Jahre auftretenden neuen Arten von Daten, Visualisierungen oder anderen Elemente mit den anfänglich aufgestellten Kategorien erfassen. Dies kann bereits als erstes Ergebnis im Hinblick auf Forschungsfrage 1 angesehen werden: Das Set an *möglichen* Komponenten des Datenjournalismus als Berichtserstattungsmuster scheint im Untersuchungszeitraum relativ stabil zu sein.

³ Die Autoren stellen das vollständige Codebuch auf Anfrage gerne zur Verfügung.

3. Ergebnisse

Die Ergebnispräsentation folgt den Forschungsfragen in umgekehrter Reihenfolge: Zunächst werden die Akteure im Datenjournalismus in den Blick genommen, d. h. die unterschiedlichen Medien und Medientypen unter den Nominierten, die Anzahl der an den Projekten beteiligten Journalisten u. Ä. Danach werden die Befunde zu behandelten Themen sowie erzählerischen Mitteln und Kritik der analysierten Fälle vorgestellt. Diese Ergebnisse liefern den Hintergrund, vor dem wir die Ergebnisse zu den eigentlichen Schlüsselvariablen des Datenjournalismus vorstellen: Datentypen und -verarbeitung, Formen der Visualisierung sowie interaktive Funktionen. Dabei wird auch auf statistisch signifikante sowie andere deutliche und inhaltlich bedeutende Unterschiede zwischen den drei Jahrgängen sowie zwischen DJA-Gewinnern und „Nur-Nominierten“ eingegangen.

3.1 Medientypen und Anzahl der beteiligten Akteure

Ein erster bemerkenswerter Befund ist, dass – zumindest unserer speziellen Stichprobe nach – Datenjournalismus als digitale Form der Berichterstattung am häufigsten und anscheinend auch am erfolgreichsten von ursprünglich nicht-digitalen Medien, nämlich von Zeitungen, betrieben wird: Über die drei analysierten Jahre hinweg stellen Tages- und Wochenzeitungen jeweils die größte Gruppe unter den Urhebern sowohl von lediglich nominierten als auch von preisgekrönten datengetriebenen Projekten (s. Tabelle 3).

Tabelle 3: Art des Mediums

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausge- zeichnet (2013- 2015) (n = 151)		Ausge- zeichnet (2013- 2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Website einer Zeitung	23	41,1	28	43,8	28	47,5	66	43,7	13	46,4	79	44,1
Website einer investiga- tiven journalistischen Or- ganisation	8	14,3	16	25,0	10	16,9	26	17,2	8	28,6	34	19,0
Website einer Zeitschrift	4	7,1	11	17,2	1	1,7	16	10,6	–	–	16	8,9
Website einer öffentlich- rechtlichen Rundfunkan- stalt	5	8,9	2	3,1	3	5,1	7	4,6	3	10,7	10	5,6
Originäres Online-Me- dium	5	8,9	2	3,1	2	3,4	8	5,3	1	3,6	9	5,0
Website einer Nachrich- tenagentur	3	5,4	–	–	6	10,2	9	6,0	–	–	9	5,0
Website einer anderen Organisation	2	3,6	2	3,1	5	8,5	9	6,0	–	–	9	5,0
Website eines Universi- tätsmediums	3	5,4	2	3,1	1	1,7	5	3,3	1	3,6	6	3,4
Website einer privaten Rundfunkanstalt	2	3,6	1	1,6	2	3,4	3	2,0	2	7,1	5	2,8
Andere	1 ^I	1,8	–	–	1 ^{II}	1,7	2	1,3	–	–	2	1,1

^I In diesem Fall: Private Website des freiberuflichen Journalisten Gregor Aisch

^{II} In diesem Fall: Kooperation zwischen einer Tageszeitung und einem monatlich erscheinenden Nachrichtenmagazin

Gestützt wird dieser Befund auch durch die Experten-Interviews, die Weinacht/Spiller (2014: 430; Hervh. im Orig.) mit 35 deutschen Datenjournalisten geführt haben:

„Die befragten Experten beschreiben den Datenjournalismus in der Summe aller Interviews als eine Dienstleistung, die überwiegend von überregionalen Printmedien nachgefragt (v. a. Stern, Spiegel, SZ, taz) oder in Blogs veröffentlicht wird. Von Radio- und TV-Sendern wird Datenjournalismus nur in Ausnahmefällen praktiziert. Dies dürfte zum einen daran liegen, dass für diese Medien die Komplexität von Sachverhalten sehr stark reduziert werden muss und dass deren Websites eher eine ergänzende als eine tragende Funktion haben.“

Die meistnominierten Print-Organisationen sind der *Guardian* (UK- und US-Ausgabe zusammengefasst) mit zwölf, die *New York Times* mit zehn, die US-Zeitschrift *Mother Jones* mit neun sowie die argentinische Zeitung *La Nación* mit acht Fällen. Die Gruppe, die am zweithäufigsten vertreten ist, stellen investigative journalistische Organisationen wie *Pro Publica* – das Medium, das mit 16 Fällen in der Summe die meisten Projekte beigetragen

hat – sowie *The Center for Public Integrity* / *The International Consortium of Investigative Journalists* (CPI/ICIJ) mit neun Fällen. Nur sieben der insgesamt 85 in den drei Jahren nominierten Medien standen jedes Jahr auf der Shortlist für den Award. Zudem sind *Pro Publica*, das CPI/ICIJ und der *Guardian* die einzigen Organisationen, die in allen drei Jahren mit mehr als einem Projekt vertreten waren. Genuine Online-Publikationen, Fernsehsender und Nachrichtenagenturen sind in deutlich geringerem Umfang vertreten.

Unter den Nominierten finden sich Projekte aus 27 verschiedenen Ländern; vierzehn Länder sind jedoch jeweils mit nur einem Projekt in den drei Jahren vertreten. Dass im Jahr 2015 acht Länder (darunter etwa Ägypten und Georgien) zum ersten Mal nominiert waren, deutet darauf hin, dass sich Datenjournalismus zunehmend zu einem weltweiten Phänomen entwickelt. Andere Länder, die bereits 2013 und/oder 2014 unter den Nominierten waren, sind 2015 allerdings nicht wieder vertreten (etwa die Niederlande, Frankreich und Ungarn). Über die gesamten drei Jahre zeigt sich eine stabile, aber in sich ebenfalls stark gestaffelte Spitzengruppe von Ländern, in der die Vereinigten Staaten mit 48,6 Prozent der Nominierungen insgesamt weit vor Großbritannien (12,8 %) und Deutschland (7,3 %) liegen. Es ist daher nicht überraschend, dass mehr als zwei Drittel (67 %) der nominierten Beiträge englischsprachig sind; Ursache hierfür könnte zudem sein, dass Datenjournalismus in englischsprachigen Ländern eine längere Tradition hat (vgl. De Maeyer et al. 2015; Parasie/Dagiral 2013; Weinacht/Spiller 2014). Am zweithäufigsten finden sich unter den Nominierten bi- oder multilinguale Projekte (15,6 %). Ihr Anteil ist über die Jahre von 14,3 auf 18,6 Prozent gestiegen und liegt unter den ausgezeichneten Arbeiten sogar bei über zwanzig Prozent (21,4 %). In den meisten dieser Fälle existieren zwei Versionen des Projekts: eine in der Originalsprache des jeweiligen Mediums und eine englischsprachige. Die häufigsten Kombinationen sind hier mit jeweils sechs bzw. fünf Fällen Englisch/Deutsch sowie Englisch/Spanisch, es finden sich aber beispielsweise auch Projekte aus Japan, Georgien, Schweden usw., die um eine englischsprachige Übersetzung ergänzt sind. Drei Beiträge existieren sogar in fünf oder mehr unterschiedlichen Sprachen. Datenjournalisten, so scheint es, wollen zunehmend auch ein internationales Publikum ansprechen.

Die Inhaltsanalyse zeigt auch, dass datenjournalistische Projekte oft in Teamarbeit entstehen. In den Fällen, in denen das Projekt Angaben zu den Beteiligten enthält ($n = 154$), werden durchschnittlich knapp über fünf Personen als Autoren oder Mitwirkende genannt ($MW = 5,16$). Fast ein Drittel (31,3 %) aller Projekte wurde zudem in Zusammenarbeit mit externen Partnern verwirklicht, die etwa zur Analyse der Daten oder der Visualisierung der Ergebnisse beigetragen haben. Allerdings ist der Anteil der Nominierungen, an denen externe Partner beteiligt waren, im Jahr 2015 wesentlich geringer als in den vorangegangenen Jahren (2013: 33,9 %; 2014: 35,9 %; 2015: 23,7 %). Dies könnte darauf hindeuten, dass Redaktionen durch Weiterbildung oder die Einstellung entsprechend qualifizierter Mitarbeiter zunehmend selbst datenjournalistische Expertise aufbauen.

Die durchschnittliche Anzahl von Personen, die an der Produktion datengetriebener Beiträge beteiligt sind, ist von etwa vier im Jahr 2013 auf knapp sechs Personen in den Jahren 2014

und 2015 gestiegen (2013: MW = 4,13, SD = 3,84; 2014: MW = 5,55, SD = 3,97; 2015: MW = 5,67, SD = 4,27). Der Unterschied ist zwar nicht statistisch signifikant (ANOVA und paarweise t-/Welch-Tests), kann aber dennoch als ein Anzeichen dafür gewertet werden, dass Datenjournalismus (auf dem Level von DJA-Nominierungen) zunehmend personalintensiv ist. Der Unterschied zwischen den Projekten, die lediglich nominiert waren (MW = 4,80, SD = 3,85, n = 129), und denen, die tatsächlich ausgezeichnet wurden (MW = 7, SD = 4,74, n = 25), ist in dieser Hinsicht größer und dazu statistisch signifikant (Welch-Test: $t = 2.186$, $df = 30.410$, $p < .05$). Der hohe Personalbedarf könnte durch die starke Arbeitsteilung in die Bereiche Recherche, Datenverarbeitung und -analyse, Visualisierung sowie Schreiben und Einordnen begründet sein, die etwa Tabary et al. (2016) sowie Weinacht/Spiller (2014) ausmachen.

3.2 In den datenjournalistischen Projekten behandelte Themen

Der Datenjournalismus in der vorliegenden Stichprobe ist stark durch politische Themen geprägt (s. Tabelle 4). Fast die Hälfte der analysierten Beiträge (48,6 %) behandelt politische Fragestellungen oder kombiniert sie mit anderen Themen. Häufig werden etwa politische und finanzielle Aspekte zusammen angesprochen, wie man etwa am Beispiel eines Projekts der Argentinischen Tageszeitung *La Nación* sieht, das detailliert die Ausgaben des Senats der vergangenen zehn Jahre auflistet.

Tabelle 4: Themen der datenjournalistischen Projekte

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausgezeichnet (2013-2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Politik	25	44,6	33	51,6	29	49,2	78	51,7	9	32,1	87	48,6
Gesellschaft	20	35,7	20	31,3	22	37,3	49	32,5	13	46,4	62	34,6
Wirtschaft	12	21,4	12	18,8	18	30,5	37	24,5	5	17,9	42	23,5
Gesundheit & Wissen- schaft	12	21,4	14	21,9	12	20,3	32	21,2	6	21,4	38	21,2
Bildung	7	12,5	2	3,1	2	3,4	11	7,3	-	-	11	6,1
Sport	2	3,6	1	1,6	3	5,1	5	3,3	1	3,6	6	3,4
Kultur	1	1,8	1	1,6	2	3,4	3	2,0	1	3,6	4	2,2

Ein typisches Thema des datengetriebenen Politik-Journalismus sind Wahlen, da diese sowohl im Vorfeld durch regelmäßige Umfragen als auch durch die endgültigen Wahlergebnisse große Mengen quantitativer Daten generieren. Beispiele hierfür sind etwa die Projekte: „Exit Polls 2012: How the Vote has Shifted“⁴, „Bundestagswahl 2013 in Berlin – Alle Stimmen der 1.709 Wahllokale“ und „Election 2015: The *Guardian* Poll Projection“. Des Weiteren wird beispielsweise die Einstellung von Politikern zu bestimmten Themen dargestellt, wie im *The Daily Beast*-Projekt „This Is Your Representative on Guns“ über die Ansichten von U.S.-Abgeordneten zu Waffengesetzen. Die costa-ricanische Zeitung *El Financiero* wiederum überprüfte mittels statistischer Daten die Behauptungen, die Präsidentschaftskandidaten in ihren Wahlkampfreden machten. Stücke wie dieses verdeutlichen, warum Datenjournalismus auch oft mit der Kontrollfunktion des Journalismus und seiner Rolle als „vierte Gewalt“ in Verbindung gebracht wird (vgl. Coddington 2015: 335; Felle 2016: 88). Der Anteil der Geschichten, die sich mit politischen Themen befassen, ist unter den ausgezeichneten Projekten allerdings geringer als unter den nur nominierten (32,1 %, n = 28 vs. 51,7 %, n = 151); dieser Unterschied ist jedoch lediglich auf dem 10 %-Niveau signifikant (Fischers exakter Test).

Gesellschaftliche Themen wie Ergebnisse von Volkszählungen oder Kriminalstatistiken sind mit über einem Drittel der Fälle (34,6 %) ebenfalls bevorzugte Themengebiete; ihr Anteil unter den ausgezeichneten Arbeiten liegt bei über 40 Prozent und damit über dem Anteil der auf Politik fokussierenden Projekte. Zahlreiche datengetriebene Projekte beschäftigen sich auch mit Themen zu dem Feld Wirtschaft (23,5 %) sowie zu Gesundheit und Wissenschaft (21,2 %; z. B. „Hooked – Canada’s Pill Problem“, „Life on the Line: 911 Breakdowns at LAFD“, „Innovative Energy Projects in Developing Countries“). Dies verdeutlicht, dass Datenjournalismus hauptsächlich jene gesellschaftlichen Bereiche in den Blick nimmt, in denen regelmäßig Daten produziert werden. Die Themen mit der geringsten Häufigkeit in unserer Stichprobe sind Bildung (6,1 %, z. B. „Därför kan Pisa-testet vara missvisande [Darum kann der Pisa-Test irreführend sein]“ oder der „Public Schools Explorer“), Sport (3,4 %) und Kultur (2,2 %; Beispiele sind „Le marché de l’art pour les nuls (Der Kunst-Markt für Dummies)“, „How to Build an Oscar Winner“ und „Front Row to Fashion Week“). Gerade im Bereich Sport überrascht der geringe Anteil (Beispiele sind „The Madness Machine“ oder „The data behind R.A. Dickey“), denn im Prinzip ist die Vielzahl im Sport geführter Statistiken geradezu prädestiniert für datenjournalistische Auswertungen.

Datengetriebene Erzählungen scheinen in der Regel einen klaren thematischen Fokus zu haben, da sich fast zwei Drittel der Fälle mit nur einer Themenkategorie befassen (65,9 %). Projekte, die zwei (29,6 %) oder drei bis vier (4,5 %) Themen miteinander verbinden, kommen deutlich seltener vor.⁵ Obgleich datenjournalistische Arbeiten häufig mehr als einen

⁴ Eine Auflistung aller nominierten Projekte der Jahre 2013, 2014 und 2015 ist verfügbar unter: http://community.globaleditorsnetwork.org/projects_by_global_event/744 [24.10.2016].

⁵ Die häufigsten Themenkombinationen waren: Politik und Gesellschaft (11,7 % der Fälle), Gesellschaft und Wirtschaft (7,8 %), Politik und Wirtschaft (7,3 %).

Datensatz (s. unten) verwenden, kombinieren sie selten Informationen aus zwei verschiedenen Themenbereichen wie beispielsweise Sport mit Gesundheit/Wissenschaft. Dies zeigt, dass sich Datenjournalismus Daten (noch) selten aus verschiedenen Perspektiven nähert oder Informationen aus unterschiedlichen sozialen Bereichen einander gegenüberstellt.

Etwas weniger als die Hälfte der datenjournalistischen Projekte nimmt Bezug auf ein einzelnes Ereignis (45,3 %) wie etwa eine Wahl, einen Kriminalfall oder ein Jubiläum. Mit einem Anteil von 57,5 Prozent tritt dies bei Arbeiten, in denen ein politisches Thema im Mittelpunkt steht, signifikant häufiger auf als bei den übrigen Nominierten (33,7 %, $n = 92$; Fishers exakter Test: $p < .01$). Im Gegensatz dazu ist der Anteil bei Berichten zu Themen aus den Bereichen Gesellschaft (32,2 %, $n = 62$), Wirtschaft (28,6 %, $n = 42$) oder Gesundheit und Wissenschaft (28,9 %, $n = 38$) signifikant geringer als bei anderen Themen (jeweils 52,1 %, $n = 117$; 50,4 %, $n = 137$; und 49,6 %, $n = 141$; Fishers exakter Test: $p < .05$). Insbesondere politischer Datenjournalismus ist also stärker auf ein einzelnes Thema fokussiert und von aktuellen Ereignissen geleitet.

3.3 Erzählerische Mittel und Elemente von Kritik

Für Datenjournalisten in Quebec stellen Tabary et al. (2016: 16) fest, dass sie die „naïve positivist position that ‚numbers speak for themselves‘“ vertreten. Der Datenjournalismus in unserem Sample hingegen verlässt sich nicht allein auf die Daten, sondern bietet oft auch zusätzliche Texte mit Hintergrundinformationen und Interpretationen an: Über 40 Prozent der Arbeiten enthalten einen begleitenden Textbeitrag (44,7 %); ein Fünftel (20,1 %) umfasst ein ganzes Dossier mit mehreren Artikeln. Allerdings verfügt auch ein Fünftel (20,7 %) der analysierten Arbeiten über keinerlei erläuternden Text, ausgenommen Überschriften und Aufmacher. Diese Projekte setzen tatsächlich allein auf die Aussagekraft numerischer Daten und ihrer Visualisierungen.

Eine Möglichkeit, die abstrakten quantitativen Daten anschaulicher zu vermitteln, ist die Einbindung eines personalisierten Fallbeispiels – auch im konventionellen Journalismus gängige Praxis. Ein Beispiel ist ein Artikel zu einem Gesundheitsthema, der mit der Geschichte eines individuellen Patienten beginnt. Eine vergleichbare Erzähltechnik haben wir in 41,9 Prozent der analysierten Arbeiten gefunden.

Datenjournalisten beschreiben sich oft als kritische Beobachter (vgl. etwa Weinacht/Spiller 2014), und tatsächlich nimmt die Hälfte der analysierten Arbeiten (49,2 %) eine kritische Position ein, beispielsweise im Hinblick auf als ungerechtfertigt verstandene Beschlagnehmungsmethoden der Polizei. In vielen dieser Fälle begnügten sich die Autoren nicht damit, Kritik an wahrgenommenen Missständen zu üben, sondern riefen die Nutzer auf, Veränderungen einzufordern – etwa im Hinblick auf mehr Sicherheit im öffentlichen Verkehrssystem –, oder verlangten selbst ein Eingreifen der Politik – beispielsweise zur Senkung der Emissionen von Kohlendioxid. Der Anteil kritischer Projekte ist über die Jahre mehr oder weniger stabil und unter den ausgezeichneten Arbeiten erheblich höher (60,7 %).

3.4 Datentypen und -verarbeitung

Die Angabe der Datenquelle gilt im Datenjournalismus als wichtiges Qualitätskriterium (vgl. Gray et al. 2012). Dennoch ist in 6,1 Prozent der Fälle in unserem Sample nicht kenntlich gemacht, woher die Daten stammen (s. Tabelle 5). Dies trifft jedoch auf keine der ausgezeichneten Arbeiten zu. Verglichen mit den Ergebnissen von Knight (2015: 65) ist dieser Anteil als gering zu betrachten, denn in den von ihr analysierten datenjournalistischen Beiträgen in nationalen britischen Zeitungen wurde in 40 Prozent der Fälle keine Datenquelle angegeben. In dieser Hinsicht können die Nominierten der Data Journalism Awards tatsächlich als „Goldstandard“ des Datenjournalismus betrachtet werden.

Mit Abstand die meisten Arbeiten in unserer Stichprobe verwenden Daten von offiziellen Institutionen wie Eurostat, den statistischen Landesämtern oder Bildungs- und Verteidigungsministerien (68,2 %); unter den ausgezeichneten Projekten beziehen sich 75 Prozent auf eine solche Quelle (s. Tabelle 5). Dieser Befund spiegelt auch die Ergebnisse von Tabary et al. (2016: 10) wider, dass Datenjournalismus – in diesem Fall in Quebec – eine „dependency on pre-processed public data“ aufweist. Politische Themen werden dabei signifikant häufiger mithilfe der Daten von offiziellen Institutionen abgedeckt (79,3 %, $n = 87$) als nicht-politische Themen (57,6 %, $n = 69$; Fishers exakter Test: $p < .01$). Die zweithäufigste Datenquelle sind andere, ebenfalls nichtkommerzielle Organisationen wie Universitäten, Forschungsinstitute und NGOs (44,1 %). Etwa 20 Prozent der Analysedaten wurden von den Medienorganisationen selbst gesammelt – beispielsweise durch eine Umfrage oder durch Recherche in den eigenen Archiven. Der Vergleich zwischen den Jahren legt nahe, dass sich Datenjournalisten zunehmend auch diesen Quellen abseits der offiziellen Institutionen zuwenden: Berichte, die sich auf Daten aus nichtkommerziellen Organisationen und privaten Unternehmen oder eigene Erhebungen beziehen, sind im Jahre 2015 sämtlich zahlreicher vertreten als noch 2013.

Diese Annahme wird durch ein weiteres Ergebnis untermauert: Über alle drei Jahre hinweg gerechnet werden zwar in mehr als der Hälfte der Fälle (mit Quellenangabe) Daten aus nur einer Quelle verwendet (53,5%); die durchschnittliche Anzahl an unterschiedlichen Datenquellen für ein datenjournalistisches Projekt ist jedoch von 1,4 Quellen im Jahr 2013 ($SD = 0.66$) auf 1,68 im Jahr 2014 ($SD = 0.63$) bzw. 1,67 im Jahr 2015 ($SD = 0.80$) gestiegen.⁶ Projekte, die mit zwei (34,9 %), drei (11,0 %) oder vier (0,6 %) unterschiedlichen Typen von Quellen arbeiten, kombinieren in den meisten Fällen offizielle Daten mit Daten entweder von anderen, ebenfalls nichtkommerziellen Organisationen (49 Fälle, 61,3 % von 80 Fällen mit zwei oder mehr Arten von Quellen), aus ihrer eigenen Organisation (21 Fälle, 26,3 %) oder von privaten Unternehmen (17 Fälle, 21,3 %).

⁶ Kruskal-Wallis-Test wegen Heteroskedastizität: $\chi^2 = 6,992$, $df = 2$, $p < .05$; paarweise Games-Howell-Tests zeigten auf dem 5 %-Niveau keine signifikanten Unterschiede und nur einen auf dem 10 %-Niveau signifikanten Unterschied zwischen den Jahren 2013 und 2014.

Tabelle 5: Art der Datenquelle (Mehrfachcodierung möglich)

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausge- zeichnet (2013- 2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Offizielle Institution	37	66,1	44	68,8	41	69,5	101	66,9	21	75,0	122	68,2
Andere, ebenfalls nicht-kommerzielle Or- ganisation	19	33,9	34	53,1	26	44,1	69	45,7	10	35,7	79	44,1
Eigene Quelle	13	23,2	9	14,1	17	28,8	32	21,2	7	25,0	39	21,8
Privates Unterneh- men	8	14,3	12	18,8	13	22,0	27	17,9	6	21,4	33	18,4
Quelle nicht ange- geben	3	5,4	5	7,8	3	5,1	11	7,3	-	-	11	6,1

In fast zwei Drittel der Projekte werden Daten verwendet, die auf nationaler Ebene erhoben wurden (65,4 %, n = 179); 19,6 Prozent beziehen sich auf internationale und 11,7 Prozent auf globale Daten. Allerdings werden im Datenjournalismus auch deutlich kleinere geographische Einheiten behandelt: Fast ein Drittel der Fälle basiert auf regionalen Daten (31,8 %), 19 Prozent analysieren Informationen auf lokaler und 7,8 Prozent gar auf hyperlokaler Ebene. Während der Anteil von Projekten, die internationale Daten nutzen, über die Jahre signifikant gestiegen ist (2013: 10,7 %, 2014: 15,6 %, 2015: 32,2 %⁷), variiert der Anteil der auf regionalen Daten basierenden Arbeiten erheblich (2013: 41,1 %, 2014: 9,4 %, 2015: 47,5 %⁸). Ein eindeutiger Trend zur „Glokalisierung“ im Datenjournalismus lässt sich hier also nicht ausmachen. Allerdings scheinen Artikel, die mit einem internationalen oder globalen Maßstab berichteten, im Wettbewerb erfolgreicher zu sein, da ihr Anteil unter den Preisträgern größer ist als unter den nur nominierten Arbeiten (25,0 % verglichen mit 14,3 %, n = 28, allerdings nicht statistisch signifikant). Eine Erklärung dafür könnte sein, dass Geschichten mit weiterem geografischen Bezug mehr soziale und sachliche Relevanz zugeschrieben wird. Doch auch Artikel, die auf hyperlokaler Ebene berichten, werden von den DJA-Juroren geschätzt: Ausgezeichnete Arbeiten basieren fast zweimal so häufig auf hyperlokalen Daten wie der allgemeine Durchschnitt (14,3 %, n = 28).

⁷ $\chi^2 = 9,412$, $df = 2$, $p < .01$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte nur einen signifikanten Unterschied zwischen den Jahren 2013 und 2015 ($p < .01$).

⁸ $\chi^2 = 23,712$, $df = 2$, $p < .001$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte zwei signifikante Unterschiede zwischen den Jahren 2013 und 2014 ($p < .001$) sowie zwischen 2014 und 2015 ($p < .001$).

Zwei Drittel der Projekte (66,5 %) benutzen Daten, die auf einfache Analyseeinheiten bezogen sind, etwa auf eine Person oder einen Flug. Komplexere Einheiten wie eine Nation oder ein Unternehmen werden in der Hälfte der Fälle verwendet (49,7 %). Nur 8,9 Prozent der Fälle thematisieren eine Analyseeinheit mittlerer Komplexität wie einen Haushalt oder eine Schulklasse.

Neben der Quelle der Daten ist ein weiterer zu berücksichtigender Aspekt die Art des Zugangs zu den Daten. Mit zwei Fünfteln stützen sich die meisten der analysierten Arbeiten auf bereits öffentlich verfügbare Daten. Etwa ein Fünftel der Projekte verwendet Daten, die bei der jeweiligen Quelle erst angefragt werden mussten. Auch auf der Grundlage des Informationsfreiheitsgesetzes beschaffte Daten sind dieser Kategorie zugeordnet – wenn sie in den Beiträgen in Infoboxen oder anderen Zusatzinformationen zu den verwendeten Daten explizit erwähnt wurden. In diesen Fällen, so eine Beobachtung während der Codierung, betonen die Datenjournalisten häufig ihre Bedeutung als Wächter in einer zu Transparenz verpflichtenden Demokratie.

In diesem Kontext überrascht es daher, dass trotz der vielzitierten Assoziation von Datenjournalismus mit Offenheit und Transparenz (z. B. Coddington 2015; Lewis/Usher 2013) in über zwei Fünftel der Fälle nicht angegeben ist, wie der Zugang zu den verwendeten Daten hergestellt wurde (s. Tabelle 6). Nach einem Rückgang 2014 steigt die Häufigkeit dieser Angaben 2015 jedoch wieder. Nur wenige Fälle basieren auf Daten, die von den Beteiligten selbst erhoben wurden (z. B. „Gastos en el Senado 2004-2013 (Senatsausgaben 2004-2013)“ und „Precios combustibles (Der Preis der Brennstoffe)“). Im Kontrast zu der erheblichen Aufmerksamkeit, welche die Berichterstattung über sogenannte „geleakte“ Daten in der jüngeren Vergangenheit erfahren hat, steht der geringe Anteil derartiger Projekte im untersuchten Sample von nur 3,4 Prozent. Insgesamt legen die geringen Anteile geleakter, angefragter und selbst erhobener Daten nahe, dass der Zusammenhang zwischen Datenjournalismus und investigativer Berichterstattung nicht so stark ist, wie häufig angenommen (vgl. Coddington 2015: 335), sondern tatsächlich eher auf wenigen, dann aber in der öffentlichen Wahrnehmung herausragenden Projekten basiert (vgl. De Maeyer et al. 2015: 433). Nichtsdestotrotz ist die Zahl der Projekte, die auf geleakten Informationen, angefragten oder selbst-generierten Daten basieren, in der vorliegenden Studie größer als in der Stichprobe von Knights (2015: 65) Untersuchung zu Datenjournalismus in überregionalen britischen Zeitungen. Auch scheint durchaus ein gewisser Zusammenhang zwischen der Art des Datenzugangs und einem journalistischen Selbstbild als Kritiker, Kontrolleur oder „vierte Gewalt“ auf: Kritik an wahrgenommenen Missständen oder sogar ein Aufruf an die Nutzer, Veränderungen einzufordern, bzw. die direkte Forderung nach einem Eingreifen der Politik finden sich signifikant häufiger in Projekten, die auf angefragten oder geleakten Daten basieren, als in Arbeiten, für welche die Daten auf andere Weise beschafft wurden.⁹ Im Gegensatz dazu

⁹ Angefragte Daten: 87,2 %, $n = 39$ vs. 38,6 %, $n = 140$, Fishers exakter Test: $p < .001$; geleakte Daten: 100,0 %, $n = 6$ vs. 47,4 %, $n = 173$, Fishers exakter Test: $p < .05$.

beinhalten Arbeiten, die öffentlich zugängliche Informationen verwenden, diese kritischen Elemente zu einem signifikant geringeren Ausmaß als solche, die sich ausschließlich auf Daten stützen, die schwieriger zu erhalten waren (37,8 %, n = 74 vs. 57,1 %, n = 105, Fishers exakter Test: $p < .05$).

Tabelle 6: Datenzugang (Mehrfachkodierung möglich)

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausgezeichnet (2013-2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Datenzugang nicht angegeben	20	35,7	28	43,8	31	52,5	70	46,4	9	32,1	79	44,1
Öffentlich verfügbare Daten	22	39,3	28	43,8	24	40,7	63	41,7	11	39,3	74	41,3
Angefragte Daten	12	21,4	10	15,6	17	28,8	29	19,2	10	35,7	39	21,8
Eigene Datensammlung	5	8,9 ¹	1	1,6 ¹	10	16,9 ¹	11	7,3	5	17,9	16	8,9
Gescrapete Data	3	5,4	5	7,8	3	5,1	10	6,6	1	3,6	11	6,1
Geleakte Daten	1	1,8	3	4,7	2	3,4	4	2,6	2	7,1	6	3,4

¹ $\chi^2 = 8,929$; $df = 2$; $p < .05$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte nur einen signifikanten Unterschied zwischen den Jahren 2014 und 2015 ($p < .01$).

Der hier untersuchte Datenjournalismus stützt sich zu einem Großteil auf Geodaten (44,4 %), Finanzdaten (43,3 %) und Messwerte, die mithilfe von Sensoren oder anderen Messwerkzeugen erhoben wurden (42,1 %; z. B. Fluglärm, Zuggeschwindigkeit, CO₂-Ausstoß, Strahlenbelastung) (s. Tabelle 7). Während derartige Messwerte über die Jahre hinweg an Bedeutung gewonnen haben, basieren die mit einem Award ausgezeichneten Projekte zu einem unterdurchschnittlichen Anteil auf derart erhobenen Daten (28,6 %). Ein anderer häufiger Datentyp sind soziodemografische Daten (32,6 %). Diese wurden signifikant häufiger in Artikeln verwendet, die sich auf eine offizielle Quelle beziehen (37,7 %, n = 122, vs. 21,1 %, n = 57; Fishers exakter Test: $p < .05$). In den ausgezeichneten Arbeiten überdurchschnittlich häufig verwendet wurden Finanz- und personenbezogene Daten, also Informationen, die individuellen Personen zugeschrieben werden können; dies ist der einzige signifikante Unterschied zwischen den Preisträgern und den nicht ausgezeichneten Arbeiten im Hinblick auf die Datenart.

Die am seltensten verwendeten Datentypen sind Befragungsergebnisse sowie Metadaten. Die letztere Kategorie umfasst „Daten über Daten“, also etwa nicht die Daten, die ein Nutzer mit einer App abgerufen oder übermittelt hat oder den Inhalt eines Telefongesprächs, wohl aber Informationen darüber, welche Apps der Nutzer wann genutzt hat bzw. wann er wie lange angerufen hat.

Wie zu erwarten werden bei bestimmten Themen einige Datenarten signifikant häufiger verwendet als andere. Beispielsweise wird auf Umfragedaten signifikant häufiger in politischen als in thematisch anders gelagerten Artikeln zurückgegriffen (23,0 %, $n = 87$, vs. 7,6 %, $n = 92$; Fishers exakter Test: $p < .01$). Zudem basieren Projekte zu Wirtschaftsthemen signifikant häufiger auf Finanzdaten (83,3 %, $n = 42$, vs. 30,7 %, $n = 137$; Fishers exakter Test: $p < .001$). Projekte zu gesellschaftlichen Themen wiederum beinhalten signifikant wahrscheinlicher soziodemographische Informationen (56,5 %, $n = 62$, vs. 19,7 %, $n = 117$; Fishers exakter Test: $p < .001$) sowie Geodaten (58,1 % vs. 36,8 %; Fishers exakter Test: $p < .01$) als die restlichen. Im Gegensatz dazu greifen solche Arbeiten zu Gesellschaftsthemen signifikant seltener auf persönliche Daten zurück (19,4 vs. 35,0 %; Fishers exakter Test: $p < .05$) als der Rest der DJA-Nominierten. Bei Themen von gesamtgesellschaftlichen Ausmaßen scheint der Perspektive von Einzelpersonen also geringere Relevanz zugeschrieben zu werden als in anderen Zusammenhängen. Messwerte wiederum werden signifikant vermehrt in Arbeiten genutzt, die (unter anderem) Wissenschaft oder Gesundheit thematisieren (78,9 %, $n = 38$, vs. 31,9 %, $n = 141$; Fishers exakter Test: $p < .001$).

Tabelle 7: Arten von Daten (Mehrfachkodierung möglich)

	2013 ($n = 55$)		2014 ($n = 64$)		2015 ($n = 59$)		Nicht ausgezeichnet (2013-2015) ($n = 151$)		Ausgezeichnet (2013-2015) ($n = 28$)		Gesamt ($n = 179$)	
		%		%		%		%		%		%
Geodaten	26	47,3	25	39,1	28	47,5	69	46,0	10	35,7	79	44,4
Finanzdaten	25	45,5	29	45,3	23	39,0	62	41,3	15	53,6	77	43,3
Messwerte	19	34,5	28	43,8	28	47,5	67	44,7	8	28,6	75	42,1
Soziodemografische Daten	21	38,2	16	25,0	21	35,6	48	32,0	10	35,7	58	32,6
Persönliche Daten	12	21,8	21	32,8	19	32,2	39	26,0 ^I	13	46,4 ^I	52	29,2
Metadaten	7	12,7	13	20,3	8	13,6	24	16,0	4	14,3	28	15,7
Befragungsdaten	8	14,5	7	10,9	12	20,3	26	17,3	1	3,6	27	15,2
Andere Daten	-	-	-	-	-	-	1 ^{II}	0,7	1 ^{III}	3,6	2	1,1

^I Fishers exakter Test: $p < .05$.^{II} Expertenberichte in „Who's Pulling the Strings of D. C. Puppet Corporations?“^{III} Gesetzestexte in „Gay Rights by State“

Die Art der Daten scheint sich auch auf den Einbezug bestimmter Inhalte und die Anwendung spezieller Erzähltechniken auszuwirken. So beinhalten auf Messwerten und/oder Finanzdaten beruhende Projekte signifikant häufiger als andere Beiträge offene Kritik an wahrgenommenen Missständen oder sogar einen Aufruf an die Nutzer, Veränderungen einzufordern, bzw. die direkte Forderung nach einem Eingreifen der Politik (60,0 %, $n = 75$, vs.

41,4 %, n = 104; Fishers exakter Test: $p < .05$; bzw. 62,3 %, n = 77, vs. 39,2 %, n = 102; Fishers exakter Test: $p < .01$).

Der Datenjournalismus in unserer Stichprobe greift jedoch selten auf nur einen Datentyp zurück: Dies ist bei lediglich rund einem Viertel der DJA-Nominierten der Fall. Die meisten Artikel basieren hingegen auf zwei oder drei verschiedenen Arten von Daten. Die durchschnittliche Anzahl unterschiedlicher Datentypen ist über die Jahre leicht angestiegen (2013: MW = 2,14; SD = 0,96; 2014: MW = 2,17; SD = 0,99; 2015: MW = 2,36, SD = 1,05). Die Chance auf eine Auszeichnung scheint aber nicht dadurch beeinflusst zu werden, wie viele unterschiedliche Datenarten in einem Projekt berücksichtigt werden (DJA-Gewinner: MW = 2,21; SD = 0,96; „Nur-Nominierte“: MW = 2,23; SD = 1,05). Am häufigsten kombiniert wurden Geodaten mit Messwerten (21,2 %; z. B. Strahlenbelastung in Becquerel oder Lärmbelastung in Dezibel an verschiedenen Orten) sowie mit soziodemografischen Informationen (17,9 %) oder Finanzdaten (16,8 %, z. B. welche Länder senden wie viel Entwicklungshilfe in welche Regionen). Darüber hinaus traten soziodemografische Statistiken häufig zusammen mit Finanzdaten (14 %) oder Messwerten (12,3 %) auf, und Finanzdaten wurden oft mit Messwerten (12,8 %) oder persönlichen Informationen (12,3 %) kombiniert.

Nahezu alle untersuchten Fälle verwendeten „statische“ Daten, die nur einmal erhoben und nicht fortgeschrieben wurden (91,1 %). Lediglich 14 Arbeiten bauen auf Daten auf, die regelmäßig aktualisiert werden (z. B. „Bloomberg Billionaires: Today’s Ranking of the World’s Richest People“), und nur zwei Projekte arbeiteten mit Echtzeitdaten: das „Tweet-o-Meter“ (Tweetómetro) von 2013 nutzte Daten direkt von der Twitter-API, und der „Influence Explorer“ (2015) aktualisiert automatisiert die Gesamtsumme und andere Daten der Wahlkampfspenden für die einzelnen Kandidaten in US-Wahlkämpfen, sobald neue Spenden auf den Regierungwebseiten dokumentiert werden.

Tabelle 8 gibt eine Übersicht darüber, worauf die Projekte bei der Analyse der Daten den Fokus legen.

Tabelle 8: Fokus der Datenanalyse (Mehrfachkodierung möglich)

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausgezeichnet (2013-2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Vergleich von Werten	46	82,1	56	87,5	56	94,9	131	86,8	27	96,4	158	88,3
Aufzeigen eines zeitlichen Verlaufs	26	46,4	30	46,9	34	57,6	75	49,7	15	53,6	90	50,3
Aufzeigen von Verbindungen und Daten-, Zahlungs- oder anderen Strömen	18	32,1	23	35,9	17	28,8	48	31,8	10	35,7	58	3,4
Aufzeigen von Hierarchien	8	14,3	6	9,4	13	22,0	26	17,2	1 ¹	3,6	27	15,1

¹ „Le marché de l'art pour les nuls (Der Kunst-Markt für Dummies)“

In der großen Mehrheit der Fälle liegt das Augenmerk der Datenanalyse auf dem Vergleichen von Werten, etwa um Unterschiede zwischen Männern und Frauen oder Nachbarschaften aufzuzeigen. Die Hälfte der Arbeiten dokumentiert Veränderungen über die Zeit (z. B. „Climate Change: How Hot Will It Get in My Lifetime?“). Verbindungen zwischen verschiedenen Akteuren sowie Daten-, Zahlungs- oder andere Ströme zwischen unterschiedlichen Stellen werden nur in etwa einem Drittel der Projekte dargestellt. Noch deutlich seltener sind Projekte, in denen mit Daten Hierarchien veranschaulicht werden. Ein Beispiel hierfür ist das Projekt „Women as Academic Authors“, in dem ein Ranking von Wissenschaftlerinnen nach ihrem Einfluss aufgestellt wird.

Bestimmte Themen scheinen einige Analyseschwerpunkte eher nahezulegen als andere: So werden Daten in Artikeln, die (auch) ein politisches Thema behandeln, signifikant häufiger dazu verwendet, Verbindungen bzw. Daten-, Zahlungs- oder andere Ströme aufzudecken, als in den anderen Projekten (43,7 %, n = 87 vs. 21,7 %, n = 92; Fishers exakter Test: $p < .01$): Beispielsweise zeigt das Projekt „Rede de Escândalos (Netzwerk der Skandale)“ Verbindungen zwischen brasilianischen Politikern und ihre Verwicklung in verschiedene politische Skandale; „Consider the Source“ folgt Zahlungsströmen von Firmen zu nichtkommerziellen Organisationen.

Die durchschnittliche Anzahl unterschiedlicher Analysezwecke pro Projekt wächst – zwar nicht signifikant, aber dennoch stetig (2013: MW = 1,75, SD = 0,75; 2014: MW = 1,80, SD = 0,76; 2015: MW = 2,03, SD = 0,81). Datenjournalismus kombiniert also vermehrt unterschiedliche Herangehensweisen und nimmt damit immer komplexere Analysen vor.

3.5 Verwendete Formen der Visualisierung

Um Datenjournalismus als spezifisches Berichterstattungsmuster beschreiben zu können, muss auch seine spezielle Art, Geschichten zu erzählen, analysiert werden. Wie im Forschungsstand dargestellt werden zu den Kernelementen insbesondere interaktive Funktionen sowie Formen der Visualisierung von Daten gezählt.

Tabelle 9 macht deutlich, dass es bei den hier analysierten Fällen ein mehr oder weniger festes Set von Visualisierungselementen gibt, das hauptsächlich aus Bildern und einfachen statischen Grafiken (jeweils 62,6 %) sowie Landkarten (48,0 %) besteht. Ein Drittel der Projekte arbeitet (auch) mit Tabellen (33,5 %). Animierte Visualisierungen sind deutlich seltener (16,8 %). Dieser Befund spiegelt zum Teil auch die Aussage der von Appelgren/Nygren (2014: 403) interviewten Datenjournalisten, die „maps as the standard visualizing method“ beschrieben, ebenso wie Knights Erkenntnis, dass die Visualisierung durch Grafiken und Landkarten in nationalen Zeitungen in Großbritannien die „most common form of data information presented“ ist (2015: 65). Der Anteil an Tabellen und Animationen ist unter den ausgezeichneten Projekten relativ hoch; jedoch ist lediglich der Prozentsatz der Abbildungen signifikant größer als bei den nur nominierten Arbeiten. Darüber hinaus hat die Verwendung von Abbildungen, Grafiken sowie Tabellen von 2013 bis 2015 signifikant zugenommen.

Schaut man auf den Zusammenhang zwischen Visualisierung und Thema des Beitrags, zeigt sich, dass Arbeiten zu gesellschaftlichen Themen häufiger Landkarten (62,9 %, $n = 62$) und animierte Visualisierungen (25,8 %, $n = 62$) beinhalten als andere Projekte (40,2 % bzw. 12,0 %, jew. $n = 117$, Fishers exakter Test: $p < .01$ bzw. $p < .05$). Darüber hinaus finden sich Animationen mit einem Anteil von acht Prozent in politischen Geschichten ($n = 87$) signifikant seltener als in Projekten ohne politischen Fokus (25,0 %, $n = 92$, Fishers exakter Test: $p < .01$).

Tabelle 9: Visualisierung (Mehrfachkodierung möglich)

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausgezeichnet (2013-2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Bilder (Foto, Zeichnung, Video)	26	46,4 ^I	46	71,9 ^I	40	67,8 ^I	89	58,9 ^{II}	23	82,1 ^{II}	112	62,6
Einfache statische Grafiken	31	55,4 ^{III}	34	53,1 ^{III}	47	79,7 ^{III}	96	63,6	16	57,1	112	62,6
Landkarten	29	51,8	30	46,9	27	45,8	75	49,7	11	39,3	86	48,0
Tabellen	14	25,0 ^{IV}	18	28,1 ^{IV}	28	47,5 ^{IV}	48	31,8	12	42,9	60	33,5
Kombinierte statische Diagramme	11	19,6	11	17,2	13	22,0	28	18,5	7	25,0	35	19,6
Animierte Visualisierungen	6	10,7	13	20,3	11	18,6	22	14,6	8	28,6	30	16,8
Andere Visualisierungen	-	-	-	-	5	8,5	5	3,3	-	-	5	2,8
Keine Visualisierungen	-	-	-	-	1 ^V	1,7	1 ^V	0,7	-	-	1 ^V	0,6

^I $\chi^2 = 9,284$; $df = 2$; $p < .01$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte zwei signifikante Unterschiede sowohl zwischen den Jahren 2013 und 2014 ($p < .01$) als auch zwischen 2013 und 2015 ($p < .05$).

^{II} Fishers exakter Test: $p < .05$.

^{III} $\chi^2 = 11,040$; $df = 2$; $p < .01$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte zwei signifikante Unterschiede sowohl zwischen den Jahren 2013 und 2014 ($p < .01$) als auch zwischen 2014 und 2015 ($p < .01$).

^{IV} $\chi^2 = 7,803$; $df = 2$; $p < .05$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigte keine signifikanten Unterschiede.

^V „Preto no branco (Schwarz in Weiß)“ analysiert die Aussagen von Politikern und bewertet diese als richtig oder falsch, ohne eine Visualisierung.

Durchschnittlich enthalten die Arbeiten mehr als zwei unterschiedliche¹⁰ Arten der Visualisierung (MW = 2,46, SD = 1,12). Oft kombiniert werden einfache statische Grafiken mit Bildmaterial (39,7 % aller Fälle) oder Landkarten (31,3 %) sowie Landkarten mit Abbildungen (28,5 %). Die Anzahl verwendeter Visualisierungselemente pro Projekt ist über die Jahre signifikant gewachsen (2013: MW = 2,09, SD = 0,92; 2014: MW = 2,38, SD = 1,13; 2015: MW = 2,90, SD = 1,16).¹¹ Dies deutet darauf hin, dass auch für die DJA-Nominierten zutrifft, was Knight (2015: 55) für die von ihr untersuchten Projekte überregionaler britischer Zeitungen feststellt: Datenjournalismus „is practiced as much for its visual appeal as for its

¹⁰ Es wurde nicht berücksichtigt, ob Elemente der gleichen Art mehr als einmal enthalten waren. Beispielsweise wurden mehrere Bilder als eine Visualisierung gezählt.

¹¹ Kruskal-Wallis-Test wegen Heteroskedastizität (Levene-Test): $\chi^2 = 16,207$; $df = 2$; $p < .001$. Der Games-Howell-Test ergab signifikante Unterschiede zwischen 2013 und 2015 ($p < .001$) sowie 2014 und 2015 ($p < .05$).

investigative qualities“. Zudem ist die durchschnittliche Zahl von Visualisierungen bei ausgezeichneten Projekten erheblich höher als bei nur nominierten (MW = 2,40, SD = 1,13, vs. MW = 2,75, SD = 1,04; der Unterschied ist aber nicht signifikant), was ebenfalls auf eine besondere Bedeutung der visuellen Elemente hinweist.

3.6 Interaktive Funktionen

Auch interaktive Funktionen werden häufig als Kernelemente des Datenjournalismus genannt (vgl. Coddington 2015; Gray et al. 2012; Weinacht/Spiller 2014). Im hier analysierten Sample finden sich allerdings in 15,1 Prozent aller Fälle keine interaktiven Elemente. Unter den ausgezeichneten Projekten ist jedoch nur ein Fall, in dem Nutzern keine Möglichkeit gegeben wird, sich die Daten interaktiv zu erschließen („Reshaping New York“, 2014, s. Tabelle 10). Dies lässt vermuten, dass Interaktivität beim DJA als Qualitätskriterium angesehen wird.

Tabelle 10: Interaktive Funktionen (Mehrfachkodierung möglich)

	2013 (n = 56)		2014 (n = 64)		2015 (n = 59)		Nicht ausgezeichnet (2013-2015) (n = 151)		Ausgezeichnet (2013-2015) (n = 28)		Gesamt (n = 179)	
		%		%		%		%		%		%
Keine interaktive Funktion	7	12,5	15	23,4	5	8,5	26	17,2 ^I	1	3,6 ^I	27	15,1
Zoom / Aufrufen von Details	32	57,1 ^{II}	35	54,7 ^{II}	46	78,0 ^{II}	95	62,9	18	64,3	113	63,1
Filter	30	53,6	32	50,0	39	66,1	85	56,3	16	57,1	101	56,4
Personalisierung	13	23,2	9	14,1	9	15,3	23	15,2	8	28,6	31	17,3
Suche	17	30,4	15	23,4	16	27,1	43	28,5	5	17,9	48	26,8
Spielerische Interaktion	2	3,6	1	1,6	3	5,1	5	3,3	1	3,6	6	3,4

^I Fishers exakter Test: $p < .05$ (einseitig).

^{II} $\chi^2 = 8,401$; $df = 2$; $p < .05$; Fishers exakter Test für paarweise Vergleiche mit angepassten α -Levels (Bonferroni-Holm-Korrektur) zeigt nur einen signifikanten Unterschied zwischen den Jahren 2014 und 2015 ($p < .01$).

Die am häufigsten verwendeten interaktiven Elemente sind Zoom-Funktionen für Landkarten, das Aufrufen von Details (etwa per Mouseover) zu einzelnen Datenpunkten (z. B. die Anzahl der Opfer für jeden Fall einer gemeldeten Schießerei in einer Schule) sowie Filterfunktionen, die es dem Benutzer ermöglichen, die bereitgestellten Daten hinsichtlich verschiedener Variablen zu filtern (z. B. um Wahlergebnisse eines bestimmten Staates oder Jahres auszuwählen). Möglichkeiten zur Personifizierung von Daten sind weniger verbreitet (17,3 % der Fälle) – in solchen Fällen muss der Nutzer persönliche Daten wie die eigene Postleitzahl oder das Alter eingeben, damit die Auswertung und Visualisierung der Daten

auf die individuellen Werte angepasst werden können. Außerdem enthielten in den drei analysierten Jahren lediglich sechs Projekte die Möglichkeit einer spielerischen Interaktion. Beispiele hierfür sind die Spiele „Heart Saver“ und „Tanastoroba“ (Entdecke deine Gehaltslücke), bei denen die Benutzer so schnell wie möglich Krankenwagen zu fiktionalen Charakteren mit einer Herzattacke schicken müssen bzw. das komplette Arbeitsleben einer Person simulieren können.

Trotz des recht hohen Anteils von Projekten ohne interaktive Möglichkeiten bietet ein durchschnittliches Projekt 1,67 unterschiedliche¹² Interaktionsfeatures (SD = 1,08); 2015 liegt dieser Wert sogar knapp unter zwei (2013: 1,68; 2014: 1,44; 2015: 1,92). Neun Projekte (5,0 %) beinhalten die maximale Anzahl von vier verschiedenen interaktiven Elementen. Am häufigsten kombiniert werden ein Zoom/Detail-Feature mit einer Filter- (42,5 % aller Fälle) oder Suchfunktion (20,7 %); Filter- und Suchfunktion werden in 20,7 % der Fälle kombiniert. Insgesamt gilt, so zeigen Kreuztabellen, dass ein Projekt, das eines dieser Features anbietet, auch signifikant eher eine der anderen beiden Funktionen enthält, als Arbeiten, in denen ersteres Feature nicht vorhanden ist.¹³ Des Weiteren zeigt sich auch ein Zusammenhang zwischen der Art des interaktiven Features und der Form der Visualisierung: Projekte, die eine Karte enthalten, bieten signifikant häufiger eine Zoom- und/oder Detailfunktion an als andere (80,2 %, $n = 86$ vs. 47,3 %, $n = 93$; Fishers exakter Test: $p < .001$). Auch in Projekten, in denen Daten in Form von Tabellen dargestellt werden, findet sich dieses Feature signifikant häufiger als im Rest der Fälle (78,3 %, $n = 60$ vs. 55,5 %, $n = 119$; Fishers exakter Test: $p < .01$). Diese Arbeiten bieten zudem mit einer höheren Wahrscheinlichkeit sowohl eine Filterungsmöglichkeit (75,0 % vs. 47,1 %; Fishers exakter Test: $p < .001$) als auch eine Suchfunktion an (38,3 % vs. 21,0 %; Fishers exakter Test: $p < .05$).

¹² Es wurde nicht berücksichtigt, ob Features der gleichen Art (z. B. Zoom-Funktion) mehr als einmal angeboten wurden (etwa in mehr als einer enthaltenen Karte).

¹³ Projekte inklusive Zoom-/Detailfunktion bieten signifikant häufiger auch eine Filterfunktion (67,3 %, $n = 113$, vs. 37,9 %, $n = 66$; Fishers exakter Test: $p < .001$) und/oder eine Suchfunktion (32,7 % vs. 16,7 %; Fishers exakter Test: $p < .05$); Projekte mit Filterfunktion bieten signifikant häufiger auch eine Zoom-/Detailfunktion (75,3 %, $n = 101$, vs. 47,4 %, $n = 78$; Fishers exakter Test: $p < .001$) und/oder eine Suchfunktion (36,6 % vs. 14,1 %; Fishers exakter Test: $p < .001$); Projekte mit Suchfunktion bieten signifikant häufiger auch eine Zoom-/Detailfunktion (77,1 %, $n = 48$, vs. 58,0 %, $n = 131$; Fishers exakter Test: $p < .05$) und/oder eine Filterfunktion (77,1 % vs. 48,9 %; Fishers exakter Test: $p < .001$).

4. Zusammenfassung und Fazit

Das „typische“ datenjournalistische Projekt weist gemäß der Vollerhebung aller Nominierungen für die Data Journalism Awards von 2013 bis 2015 folgende Eigenschaften auf: Es...

- wurde auf der Website einer Zeitung veröffentlicht,
- wurde in den USA produziert,
- ist in englischer Sprache verfasst,
- ist das Werk eines Autorenteam,
- behandelt ein politisches Thema,
- enthält mindestens einen dazugehörigen Textbeitrag,
- bezieht sich auf öffentlich zugängliche Daten aus einer offiziellen Quelle,
- basiert auf vorzugsweise auf nationaler Ebene erhobenen Finanzinformationen, Geodaten und/oder Messwerten, die sich auf eine einfache Analyseeinheit, etwa einzelne Personen, beziehen,
- vergleicht Werte, um Unterschiede und Gemeinsamkeiten darzustellen (z. B. zwischen Personen unterschiedlichen Geschlechts, Nachbarschaften) und/oder zeigt Veränderungen über die Zeit,
- bietet zwei Arten von Visualisierungen – vorzugsweise Bilder kombiniert mit Karten oder einfachen Tabellen oder Grafiken –
- und ermöglicht dem Benutzer, in eine Karte zu zoomen, Detailinformationen abzufragen und/oder Daten nach vorgegebenen Kategorien zu filtern.

Dies zeigt, dass Datenjournalismus als Berichterstattungsmuster deutlich von den strukturellen Elementen gekennzeichnet ist, die sich in den bisherigen Beobachtungen, der Literaturrecherche und den Akteursstudien bereits als Schlüsselmerkmale abgezeichnet haben.

Allerdings stellt diese Liste eine stark vereinfachte Übersicht dar, da die aufgeführten Charakteristika nicht auf alle datenjournalistischen Projekte zutreffen und es eine Vielzahl anderer Erzählelemente gibt, die zwar weniger häufig (in verschiedenen Kombinationen) verwendet werden, jedoch häufig genug, um von Bedeutung zu sein. Dies verdeutlicht auch, dass Datenjournalismus als Berichterstattungsmuster noch in der Entwicklung begriffen ist und verschiedene Datentypen, Analyse- und Visualisierungsstrategien je nach Thema und zu erzählender Geschichte flexibel kombiniert werden. Tatsächlich finden sich einige „typische Kombinationen“ von Themen, Datentypen und bestimmten Visualisierungen: Beispielsweise basieren politische Geschichten signifikant häufiger auf Befragungen als Projekte zu anderen Themengebieten, während Wirtschaftsthemen naheliegender Weise häufig auf der Grundlage von Finanzdaten behandelt werden; gesellschaftliche Aspekte werden mit Geo- und soziodemografischen Daten erfasst und Artikel zu Gesundheit und Wissenschaft beziehen sich oft auf Messwerte. Geo- und soziodemografische Daten werden signifikant häufiger durch Karten dargestellt als andere Arten von Informationen, personenbezogene Daten hingegen häufiger von Bildern begleitet und finanzielle Informationen in Tabellen aufbereitet.

Auch zwischen den „nur“ nominierten und den tatsächlich ausgezeichneten Projekten werden Unterschiede deutlich: So sind Projekte von privaten oder öffentlichen Rundfunkanstalten sowie insbesondere von investigativen Organisationen überdurchschnittlich häufig unter den Preisträgern vertreten. Im Gegensatz dazu wurden Projekte von Printmagazinen, nicht primär journalistischen Organisationen (etwa Universitäten und Forschungseinrichtungen) sowie Nachrichtenagenturen bisher nur nominiert, jedoch nie prämiert.

Ausgezeichnete Geschichten sind – im Gegensatz zu den lediglich nominierten – häufiger durch folgende Merkmale charakterisiert: Sie...

- entstehen unter Mitarbeit einer größeren Anzahl von Autoren (sind also personalintensiver),
- decken eher Themen ab, die typischerweise dem Gesellschaftsressort zugeordnet werden, wohingegen Politik und Wirtschaft weniger häufig Gegenstand sind,
- stehen weniger häufig im Zusammenhang mit einem bestimmten Ereignis, sondern behandeln eher Aspekte von längerfristiger Relevanz,
- werden in zwei oder mehr unterschiedlichen Sprachen veröffentlicht,
- werden in Kooperation mit externen Partnern produziert,
- nennen eher die Quelle der verwendeten Daten, und diese Quelle ist häufiger eine offizielle Institution und seltener eine andere, ebenfalls nichtkommerzielle Organisation,
- geben eher an, wie sie an die Daten gelangt sind, und gewinnen ihre Daten danach auch öfter durch Anfragen (etwa auf Grundlage des Informationsfreiheitsgesetzes), Leaks und/oder eigene Erhebungen,
- verwenden Finanz- und personenbezogene Daten als Grundlage, wohingegen Geodaten, Messwerte sowie Ergebnisse von Umfragen seltener genutzt werden,
- nehmen Bezug auf einen globalen, internationalen oder hyperlokalen Kontext sowie auf eine einfache Analyseeinheit,
- verwenden ein personalisiertes Fallbeispiel,
- enthalten offene Kritik oder einen Aufruf zu öffentlichem Eingreifen und nehmen damit auch eine Haltung als „Watchdog“ ein,
- vergleichen Werte, während Hierarchien seltener aufgezeigt werden,
- präsentieren Daten in Form einer Tabelle, einem kombinierten statischen Diagramm, einem Bild oder einer Animation, wohingegen weniger häufig einfache statische Grafiken oder Karten verwendet werden,
- ermöglichen eine Personalisierung der Daten, wobei Suchfunktionen oder fehlende interaktive Features seltener vorkommen.

Diese Beobachtungen sind jedoch lediglich als Tendenzen zu betrachten, da die Unterschiede zwischen den ausgezeichneten und den nur nominierten Projekten in den wenigsten Fällen statistisch signifikant sind. Die Ausnahme sind die Unterschiede in Bezug auf die Verwendung von Bildern als Visualisierung, personenbezogenen Daten und interaktiven Features (auf dem 5 %-Niveau) sowie die Häufigkeit von Politik als Themenbereich und Daten aus Befragungen (auf dem 10 %-Niveau). Datengetriebene Arbeiten bekommen also

nicht nach dem Prinzip „je mehr, desto besser“ einen Data Journalism Award verliehen – sei es in Bezug auf die Anzahl behandelter Themen, genutzter Datentypen oder -quellen, verwendeter Visualisierungen oder angebotener interaktiver Funktionen. Datenanalyse und -visualisierung soll hier demnach nicht zum Selbstzweck eingesetzt werden, sondern die erzählte Geschichte unterstützen. Dieses Ergebnis deckt sich auch mit Coddingtons Beobachtung, dass Datenjournalisten die Verwendung von Daten im Allgemeinen dem „professional journalistic value of narrative and the ‚story‘“ (2015: 339) unterordnen.

Zu berücksichtigen ist bei alledem, dass das hier analysierte Sample Einschränkungen unterliegt, da es auf Fällen basiert, die eine zweifache Verzerrung aufweisen: Erstens repräsentieren sie als Nominierte eines Awards für Datenjournalismus eine spezielle Gruppe, die nicht mit einem „alltäglichen“ Datenjournalismus gleichgesetzt werden kann. Zweitens basieren diese Projekte auf einer Selbstauswahl, da sie von den jeweiligen Medien oder Autoren selbst eingereicht werden müssen, um für eine Nominierung und mögliche Auszeichnung durch die Jury in Betracht zu kommen. Neben diesen Einschränkungen hat unsere Stichprobe aber auch zwei spezielle Vorzüge: Erstens können wir annehmen, dass die analysierten Fälle als Nominierte für den DJA einen gewissen Qualitätsanspruch erfüllen und dass insbesondere die ausgezeichneten Werke von der Jury, die als Experten des Feldes betrachtet werden können, als „Goldstandard“ angesehen werden. Als solcher können sie durchaus die zukünftige Entwicklung des Datenjournalismus beeinflussen. Zweitens ermöglicht der Vergleich von drei aufeinander folgenden Jahren, die Entwicklung des Feldes nachzuzeichnen. Am auffälligsten ist, dass...

- es lediglich wenige Organisationen schafften, jedes Jahr unter den Nominierten zu sein,
- die Herkunft der Nominierten sich im Laufe der Jahre auf immer mehr Länder verteilt hat, Datenjournalismus also ein zunehmend globales Phänomen zu werden scheint,
- Datenjournalismus die Internationalisierung journalistischer Berichterstattung und ihre globale Verbreitung zu fördern scheint, da ein beträchtlicher Teil der Arbeiten in zwei oder mehr Sprachen veröffentlicht wurde,
- die Produktion von Datenjournalismus (auf dem Wettbewerbsniveau) ein zunehmend personalintensives Unterfangen ist,
- Datenjournalisten neben den offiziellen und öffentlich zugänglichen Datenquellen vermehrt auch andere Quellen erschließen,
- datengetriebene Arbeiten zunehmend verschiedene Arten von Datenquellen sowie verschiedene Datentypen kombinieren und sich vermehrt auf internationale Daten stützen,
- die durchschnittliche Anzahl an visuellen Elementen, die in einem Projekt kombiniert werden, über die Jahre gestiegen ist.

Alles in allem können diese Indikatoren als Hinweise darauf interpretiert werden, dass datenjournalistische Projekte komplexer werden – zumindest was die Nominierten des DJA betrifft. Gleichzeitig trifft aber auch das Gegenteil zu, denn der „alltägliche“ Datenjournalismus wird zunehmend einfacher zu produzieren, da immer mehr Tools und Applikationen

entwickelt und veröffentlicht werden, die den Einstieg erleichtern (wie z. B. der „Datawrapper“: <https://datawrapper.de/>). Darüber hinaus kann der beträchtliche Rückgang beim Einsatz von externen Partnern dahingehend interpretiert werden, dass offenbar vermehrt auch innerhalb von Redaktionen die notwendige Expertise aufgebaut wird.

Dass hier nicht der „alltägliche“ Datenjournalismus Gegenstand der Analyse ist, verdeutlicht auch der Anteil investigativer Projekte, die sich auf geleakte sowie angefragte Daten beziehen und/oder von Organisationen wie *Pro Publica* veröffentlicht wurden, die sich speziell dieser Form des Journalismus widmen: Zwar sind solche Fälle immer noch seltener als Projekte, die sich auf ohnehin öffentlich zugängliche Daten stützen; dennoch sind sie zahlreich vertreten. Im Gegensatz dazu stellt Knight (2015: 69) bei ihrer Analyse von Datenjournalismus in überregionalen britischen Zeitungen fest, dass „Freedom of Information actions and leaks were not widely represented“.

Coddington (2015: 335) sieht Datenjournalismus ebenfalls als „much more loosely coupled with investigative journalism than was CAR [computer-assisted reporting]“, obwohl er häufig im Kontext von investigativen Projekten zum Einsatz käme. Letzteres spiegelt sich auch in der organisatorischen Einordnung von Datenjournalisten wider: Beispielsweise stellten Fink/Anderson (2015: 336) in ihren Interviews mit 23 Datenjournalisten von US-Zeitungen und Online-Nachrichtenseiten fest, dass einige von ihnen Investigativ-Teams in den Redaktionen zugeordnet sind, während andere der Online-, IT- oder Grafikabteilung angehören. Dies sind exakt die drei Eckpfeiler, zwischen denen Datenjournalismus häufig verortet wird: Ein Schwerpunkt liegt auf (investigativer) Datenrecherche, einer auf der statistischen Analyse und ein weiterer auf Visualisierung und Grafik- bzw. Webdesign (vgl. Anderson 2013: 1007; Coddington 2015: 335; Royal 2012), wobei alle drei Aspekte häufig Programmierkenntnisse voraussetzen.

In der vorliegenden Studie wurde außerdem deutlich, dass die Jury der Data Journalism Awards bevorzugt datengetriebenen Journalismus mit einer kritischen Haltung und einem Selbstverständnis als „Watchdog“ auszeichnet. Bisher basieren derart kritische Projekte zum überwiegenden Teil auf bereits öffentlich zugänglichen Daten. Man kann daher nur vermuten, welches Potential ein kritischer Datenjournalismus entwickeln kann, wenn er zunehmend auch nicht öffentlich zugängliche Daten erschließt. Einige Projekte z. B. vom *Guardian*, *The New York Times* und *Pro Publica* haben dieses Potential bereits gezeigt (vgl. Coddington 2015: 335; Parasie 2015: 364). Vor diesem Hintergrund schlägt der Reporter und Dozent für Datenjournalismus Jonathan Stoneman (2015: o. S.) vor: „[W]hile Open Data is being touted by governments as their being open and transparent, journalists should be tracking what is really happening, what data are not being released, and why“. Diese Aussage macht auch deutlich, dass die Entstehung eines Phänomens wie Datenjournalismus auch die Herausbildung eines neuen Zweigs von Öffentlichkeitsarbeit nach sich zieht: dem „data-spin“, der datengetriebene Berichterstattung dadurch in seinem Sinne zu beeinflussen sucht, dass bestimmte Daten zu bestimmten Zeitpunkten auf bestimmte Weise veröffentlicht oder

eben zurückgehalten werden. Die absehbare Ko-Entwicklung und wechselseitige Beeinflussung mit einem in diesem Zuge entstehenden Feld der „Daten-Public Relations“ dürfte die (R)Evolution des datenjournalistischen Berichterstattungsmusters in Zukunft weiter dynamisieren.

Literatur

- Anderson, Chris W. (2013): Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7), S. 1005–1021.
- Appelgren, Ester; Nygren, Gunnar (2014): Data journalism in Sweden. Introducing new methods and genres of journalism into „old“ organizations. *Digital Journalism*, 2(3), S. 394–405.
- Beer, David (2009): Power through algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11(6), S. 985–1002.
- Borges-Rey, Eddy (2016): Unravelling data journalism. A study of data journalism practice in British newsrooms. *Journalism Practice*, 10(7), S. 833–843.
- Boyles, Jan Lauren; Meyer, Eric (2016): Letting the data speak. Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism*, 4(7), S. 944–954.
- Coddington, Mark (2015): Clarifying journalism's quantitative turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), S. 331–348.
- De Maeyer, Juliette; Libert, Manon; Domingo, David; Heinderyckx, François; Le Cam, Florence (2015): Waiting for data journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism*, 3(3), S. 432–446.
- Felle, Tom (2016): Digital watchdogs? Data reporting and the news media's traditional "fourth estate" function. In: *Journalism*, 17(1), S. 85–96.
- Filipović, Alexander (2015): Die Datafizierung der Welt. Eine ethische Vermessung des digitalen Wandels. *Communicatio Socialis*, 48(1), S. 6–15.
- Fink, Katherine; Anderson, Chris W. (2015): Data journalism in the United States. Beyond the „usual suspects“. *Journalism Studies*, 6(4), S. 467–481.
- Gray, Jonathan; Bounegru, Liliana; Chambers, Lucy (2012): *The data journalism handbook. How journalists can use data to improve the news.* (Early release). Sebastopol: O'Reilly.
- Holtermann, Hannes (2011): *Datenjournalismus: eine neue Form der journalistischen Wertschöpfung aus Daten.* Unveröff. Masterarbeit. Hamburg.
- Karlsen, Joakim; Stavelin, Eirik (2014): Computational journalism in Norwegian newsrooms. *Journalism Practice*, 8(1), S. 34–48.
- Knight, Megan (2015): Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16(1), S. 55–72.
- Krippendorff, Klaus (2013): *Content analysis: an introduction to its methodology.* Los Angeles: SAGE.
- Lanosga, Gerry (2014): New views of investigative reporting in the twentieth century. *American Journalism*, 31(4), S. 490–506.
- Lewis, Seth C. (2015): Journalism in an era of big data. *Digital Journalism*, 3(3), S. 321–330.
- Lewis, Seth C.; Usher, Nikki (2013): Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, 35(5), S. 602–619.
- Lewis, Seth C.; Usher, Nikki (2014): Code, collaboration, and the future of journalism. A case study of the Hacks/Hackers global network. *Digital Journalism*, 2(3), S. 383–393.
- Lombard, Matthew; Snyder-Duch, Jennifer; Bracken, Cheryl Campanella (2002): Content Analysis in Mass Communication. Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, 28(4), S. 587–604.
- Mayer-Schönberger, Viktor; Kenneth Cukier (2013): *Big Data. Die Revolution, die unser Leben verändern wird.* München: Redline.
- Parasie, Sylvain (2015): Data-driven revelation? Epistemological tensions in investigative journalism in the age of „big data“. *Digital Journalism*, 3(3), S. 364–380.
- Parasie, Sylvain; Dagiral, Eric (2013): Data-driven journalism and the public good. „Computer-assisted-reporters“ and „programmer-journalists“ in Chicago. *New Media & Society*, 15(6), S. 853–871.
- Royal, Cindy (2012): The journalist as programmer: a case study of the New York Times Interactive News Technology department. *#ISOJ: the Official Research Journal of the International Symposium for Online Journalism*, 2(1), S. 5–24.
- Royal, Cindy; Blasingame, Dale (2015): Data journalism: An explication. *#ISOJ: the Official Research Journal of the International Symposium for Online Journalism*, 5(1), S. 24–46.
- Schmidt, Siegfried J.; Weischenberg, Siegfried (1994): Mediengattungen, Berichterstattungsmuster, Darstellungsformen. In: Merten, Klaus; Schmidt, Siegfried J.; Weischenberg, Siegfried (Hrsg.): *Die Wirklichkeit der Medien.* Opladen: Westdeutscher Verlag, S. 212–236.

- Splendore, Sergio; Di Salvo, Philip; Eberwein, Tobias; Groenhardt, Harmen; Kus, Michael; Porlezza, Colin (2016): Educational strategies in data journalism: a comparative study of six European countries. *Journalism*, 17(1), S. 138–152.
- Stoneman, Jonathan (2015): „Open data“ + journalism = ? In: *Data Driven Journalism. Where Journalism Meets Data*. URL: http://datadrivenjournalism.net/news_and_analysis/open_data_journalism#sthash.3s5QTVt6.dpuf [28.10.2016].
- Tabary, Constance; Provost, Anne-Marie; Trottier, Alexandre (2016): Data journalism's actors, practices and skills: A case study from Quebec. *Journalism: Theory, Practice, and Criticism*, 17(1), S. 66–84.
- Tandoc Jr., Edson C.; Oh, Soo-Kwang (2015): Small Departures, Big Continuities? Norms, values, and routines in *The Guardian's* big data journalism. *Journalism Studies*, DOI: 10.1080/1461670X.2015.1104260.
- Wahl-Jorgensen, Karin (2013a): Subjectivity and story-telling in journalism. Examining expressions of affect, judgement and appreciation in Pulitzer Prize-winning stories. *Journalism Studies*, 14(3), S. 305–320.
- Wahl-Jorgensen, Karin (2013b): The strategic ritual of emotionality: a case study of Pulitzer Prize-winning articles. *Journalism*, 14(1), S. 129–145.
- Weinacht, Stefan; Spiller, Ralf (2014): Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik*, 59(4), S. 411–433.